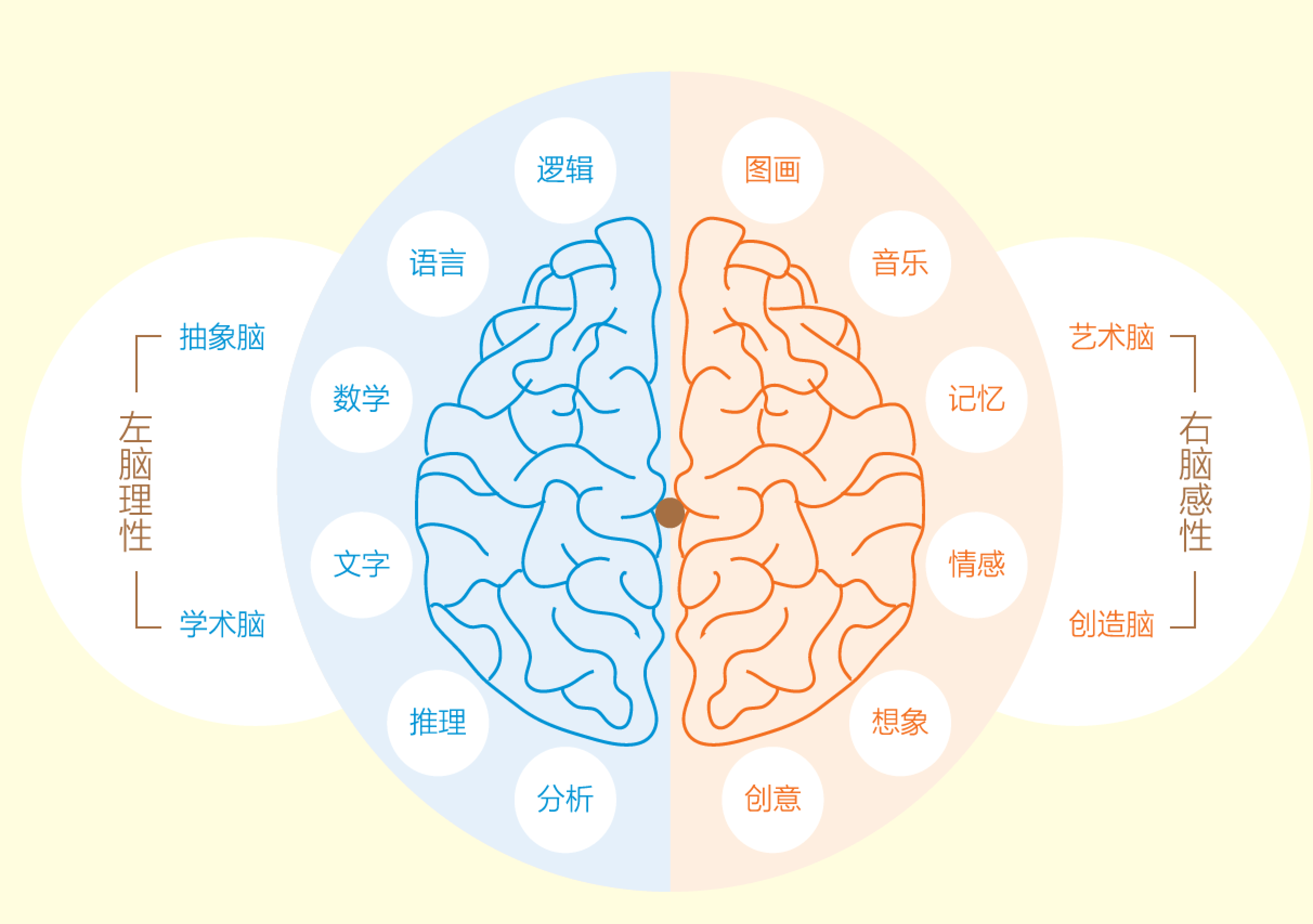
Intelligence Science

# 智能科学

## （第3版）

史忠植 著




清华大学出版社

# INTELLIGENCE SCIENCE

# 智能科学（第 3 版）

史忠植　著

清华大学出版社
北　京

## 内 容 简 介

智能科学研究智能的本质和实现技术，是由脑科学、认知科学、人工智能等创建的前沿交叉学科。脑科学从分子水平、细胞水平、行为水平研究自然智能机理，建立脑模型，揭示人脑的本质；认知科学是研究人类感知、学习、记忆、思维、意识等人脑与心智活动过程的科学；人工智能研究用人工的方法和技术，模仿、延伸和扩展人的智能，实现机器智能。智能科学不仅要进行功能仿真，而且要从机理上研究和探索智能的新概念、新理论、新方法。

本书系统地介绍智能科学的概念和方法，吸收了脑科学、认知科学、人工智能、信息科学、形式系统、哲学等方面的研究成果，探索自然智能和机器智能的机理与规律。

本书可作为大学高年级本科生和研究生的"智能科学""认知科学""神经信息学"等课程的教科书，也可作为从事智能科学、人工智能、认知科学、脑科学、神经科学、心理学等领域的研究人员的参考书。



图书在版编目(CIP)数据

智能科学/史忠植著. —3 版. —北京：清华大学出版社，2019
ISBN 978-7-302-51549-4

Ⅰ. ①智…　Ⅱ. ①史…　Ⅲ. ①人工智能　Ⅳ. ①TP18

中国版本图书馆 CIP 数据核字(2018)第 249554 号

责任编辑：王一玲
封面设计：傅瑞学
责任校对：梁　毅
责任印制：沈　露

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 刷 者：三河市铭诚印务有限公司
装 订 者：三河市启晨纸制品加工有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　印　张：34.75　　字　　数：842 千字
版　　次：2007 年 10 月第 1 版　2019 年 7 月第 3 版　　印　　次：2019 年 7 月第 1 次印刷
定　　价：89.00 元

---

产品编号：078017-01

# 前言

>>

智能科学研究智能的本质和实现技术，是由脑科学、认知科学、人工智能等创建的前沿交叉学科。脑科学从分子水平、细胞水平、行为水平研究自然智能机理，建立脑模型，揭示人脑的本质；认知科学是研究人类感知、学习、记忆、思维、意识等人脑与心智活动过程的科学；人工智能研究用人工的方法和技术，模仿、延伸和扩展人的智能，实现机器智能。智能科学不仅要进行功能仿真，而且要从机理上研究和探索智能的新概念、新理论、新方法。智能的研究不仅要运用推理，自顶向下，而且要通过学习，由底向上，两者并存。智能科学运用综合集成的方法，对开放系统的智能性质和行为进行研究。

智能科学是生命科学的精华、信息科学技术的核心，现代科学技术的前沿和制高点，涉及自然科学的深层奥秘，触及哲学的基本命题。因此，在智能科学上一旦取得突破，将对国民经济、社会进步、国家安全产生深刻而巨大的影响。目前，智能科学正处在方法论的转变期、理论创新的高潮期和大规模应用的开创期，充满原创性机遇。

智能科学的兴起和发展标志着对以人类为中心的认知和智能活动的研究已进入到新的阶段。智能科学的研究将使人类自我了解和自我控制，把人的知识和智能提高到空前未有的高度。生命现象错综复杂，许多问题还没有得到很好的说明，而能从中学习的内容也是大量的、多方面的。如何从中提炼出最重要的、关键性的问题和相应的技术，这是许多科学家长期以来追求的目标。要解决人类在21世纪所面临的许多困难，诸如能源的大量需求、环境的污染、资源的耗竭、人口的膨胀等，单靠现有的科学成就是很不够的。必须向生物学习，寻找新的科技发展的道路。智能科学的研究将为智能革命、知识革命和信息革命建立理论基础，为智能系统的研制提供新概念、新思想、新途径。

进入21世纪以来，国际上对智能科学及其相关学科，诸如脑科学、神经科学、认知科学、人工智能的研究高度重视。2013年1月28日，欧盟启动了旗舰“人类大脑计划”(human brain project)。2013年4月2日，美国启动BRAIN计划。我国也在积极筹备“脑科学与类脑研究计划”。为了争夺高科技的制高点，国务院于2017年7月8日正式发布《新一代人工智能发展规划》，力图在新一轮国际科技竞争中掌握主导权。

本书系统地介绍智能科学的概念和方法，吸收了脑科学、认知科学、人工智能、信息科学、形式系统、哲学等方面的研究成果，综合地探索人类智能和机器智能的性质和规律。2006年出版第1版、2013年出版第2版以来，国内外在该领域的研究取得了极大进展，我们也取得了不少成果。为了反映智能科学的最新研究成果和发展方向，对原书第2版作了全面修改，特别增加了认知结构、类脑智能等内容。全书共分16章。第1章是绪论，介绍智能科学兴起的科学背景和研究内容。第2章介绍智能科学的生理基础。第3章讨论神经计算的进展。第4章探讨重要的心智模型。第5章论述视觉感知理论。第6章讨论听觉信息处

理；语言的发展对人类大脑的进化发生重大影响。第 7 章讨论语言认知的理论。第 8 章重点论述重要的学习理论和方法；记忆是思维的基础。第 9 章探讨记忆机制。第 10 章重点讨论思维形式和类型。第 11 章研究智力的发展。第 12 章讨论情绪和情感的有关理论。第 13 章初步探讨意识问题；认知结构是智能科学的重要理论基础。第 14 章讨论认知结构。第 15 章介绍智能机器人研究的进展。第 16 章介绍大数据智能和认知计算，概述国际上重大的类脑智能计划的研究进展和基本原理，展望智能科学发展路线图。

在本书撰写过程中，作者与美国麻省理工学院(MIT)明斯基(Marvin Minsky)教授、加利福尼亚大学伯克利分校扎德(Lotfi A. Zadeh)教授、斯坦福大学心智与脑计算中心麦克伦特(J. L. McClelland)教授、华盛顿大学圣路易斯分校范埃森(David Van Essen)教授、南加州大学罗森勃卢姆(P. S. Rosenbloom)教授、密歇根大学莱尔德(J. E. Laird)教授、卡内基梅隆大学米切尔(T. M. Mitchell)教授、西北大学福伯斯(K. D. Forbus)教授、密歇根州立大学翁巨扬(J. Weng)教授、加拿大滑铁卢大学伊莱亚史密斯(C. Eliasmith)教授、德国海德堡大学迈耶(K. Meier)教授、德累斯顿工业大学巴德尔(F. Baader)教授等的讨论和交流，对本书学术思想的确立和发展发挥了重要作用，在此谨向上述学者表示衷心的感谢。

本书研究工作得到国家重点基础研究发展计划课题“脑机协同的认知计算模型”(No. 2013CB329502)、“非结构化信息(图像)的内容理解与语义表征”(No. 2007CB311004)；自然科学基金重点项目“基于云计算的海量数据挖掘”(No. 61035003)、“基于感知学习和语言认知的智能计算模型研究”(No. 60435010)、“Web 搜索与挖掘的新理论与方法”(No. 60933004)等的支持；国家 863 高技术项目“海量 Web 数据内容管理、分析挖掘技术与大型示范应用”(No. 2012AA011003)“软件自治愈与自恢复技术”(No. 2007AA01Z132)等项目的支持；清华大学出版社对本书的出版给予了大力支持，在此一并致谢。

本书可作为大学高年级和研究生的“智能科学”“认知科学”“认知信息学”“人工智能”等课程的教科书，对从事智能科学、脑科学、认知科学、人工智能、神经科学、心理学、哲学等领域的研究人员也具有重要的参考价值。

智能科学是处于研究发展中的前沿交叉学科，许多概念和理论尚待探讨，加之作者水平有限，撰写时间仓促，因此书中难免存在错误或不妥之处，恳请读者指正。

史忠植

2018 年 1 月于北京

# 目 录

>>

# 第1章<<
# 绪论

智能科学研究智能的基本理论和实现技术，是由脑科学、认知科学、人工智能等创建的前沿交叉学科[923]。脑科学从分子水平、细胞水平、行为水平研究人脑智能机理，建立脑模型，揭示人脑的本质。认知科学是研究人类感知、学习、记忆、思维、意识等人脑心智活动过程的科学。人工智能研究用人工的方法和技术，模仿、延伸和扩展人的智能，实现机器智能。智能科学是实现人类水平的人工智能的重要途径。

## 1.1 智能革命

工具制造、农业革命、工业革命是人类历史上具有重大影响的三次革命。这些革命使社会、经济、文明的情况发生了重大变化，从一种方式转变到另一种方式。工业革命发生在18世纪中叶，英国人瓦特改良了蒸汽机，引起了从手工劳动向动力机器生产转变的重大飞跃，以机器取代人力，以大规模工厂化生产取代个体工场手工生产，制造、矿山、交通等产业发生巨大变化，极大地推动了社会经济和文化的发展。

工业革命让机器代替了人类的体力劳动，带来经济和社会的进步。人类一直在不懈的努力，使机器能代替人类智力劳动。

亚里士多德(Aristotle，公元前384—前322)在《工具论》中提出形式逻辑。培根(F. Bacon，1561—1626)在《新工具》中提出归纳法。莱布尼茨(G. W. von Leibnitz，1646—1716)研制了四则计算器，提出了“通用符号”和“推理计算”的概念，使形式逻辑符号化，这是“机器思维”研究的萌芽。

19世纪以来，数理逻辑、自动机理论、控制论、信息论、仿生学、计算机、心理学等学科的进展，为人工智能的诞生提供了思想、理论和物质基础。布尔(G Boole，1815—1864)创立了布尔代数，他在《思维法则》一书中，首次用符号语言描述了思维活动的基本推理法则。哥德尔(K. Godel，1906—1978)提出了不完备性定理。1936年，图灵(A M Turing，1912—1954)提出了理想计算机模型——图灵机，以离散量的递归函数作为智能描述的数学基础，创立了自动机理论。1943年，心理学家麦克洛奇(W. S. McCulloch)和数理逻辑学家皮兹(W. Pitts)在《数学生物物理公报》(*Bulletin of Mathematical Biophysics*)上发表了关于神经网络的数学模型，提出了MP神经网络模型，开创了人工神经网络的研究。1945年，冯·诺依曼(John von Neumann)提出了存储程序概念。1946年，埃克特(J. P. Eckert)和莫奇利(J. W. Manochly)研制成功ENIAC电子数字计算机。1948年香农(C. E. Shannon)发表了《通讯的数学理论》，标志一门新学科——信息论的诞生。同年，维纳(N. Wiener)创立了控制论。

中国曾经发明了不少智能工具和机器。例如，算盘是应用广泛的古典计算机；水运仪

象台是天文观测与星象分析仪器;候风地动仪是测报与显示地震的仪器。我们祖先提出的阴阳学说蕴涵着丰富的哲理,对现代逻辑的发展有重大影响。

1956 年夏天,美国达特茅斯(Dartmouth)大学的青年助教麦卡锡(John McCarthy)、哈佛大学明斯基(M Minsky)、贝尔实验室香农和 IBM 公司信息研究中心罗彻斯特(N Lochester)发起召开了达特茅斯会议。他们邀请了卡内基梅隆大学纽厄尔(A Newell)和西蒙(H A Simon)、麻省理工学院塞夫里奇(O Selfridge)和索罗门夫(R Solomamff),以及 IBM 公司塞缪尔(A Samuel)和莫尔(T More)参加。他们的研究专业包括数学、心理学、神经生理学、信息论和计算机科学,多学科交叉,从不同的角度共同探讨人工智能的可能性。麦卡锡在 *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* 中首先引入了人工智能(Artificial Intelligence,AI)术语,他将人工智能定义为:使一部机器的反应方式就像是一个人在行动时所依据的智能。达特茅斯会议标志着人工智能的正式诞生。

60 多年来,人工智能学者提出的启发式搜索、非单调推理丰富了问题求解的方法。大数据、深度学习、知识发现等的研究推动了智能系统的发展,取得实际效益。模式识别的进展,已经在一定程度上使计算机具备了听、说、读、看的能力。

2011 年 2 月 14 日至 16 日,IBM 人工智能系统“沃森”在美国著名智力竞答电视节目《危险边缘》(*Jeopardy*)中,战胜了两名“常胜将军”詹宁斯和鲁特尔。2016 年 3 月 9 日至 15 日,谷歌 AlphaGo 采用深度强化学习和蒙特卡罗搜索算法,以 4∶1 战胜了围棋冠军韩国棋手李世石。2017 年 12 月 5 日,谷歌 DeepMind 团队的西尔弗(David Silver)与哈萨比斯(Demis Hassabis)等发表论文,介绍通用棋类 AI AlphaZero,从零开始训练,除了基本规则没有任何其他知识,8 小时击败李世石版 AlphaGo,训练 34 小时的 AlphaZero 胜过了训练 72 小时的 AlphaGo Zero[718]。这些重大事件标志着“智能革命”时代已经到来。

为了争夺高科技的制高点,国务院于 2017 年 7 月 8 日正式发布《新一代人工智能发展规划》。规划指出:“人工智能成为国际竞争的新焦点。人工智能是引领未来的战略性技术,世界主要发达国家把发展人工智能作为提升国家竞争力、维护国家安全的重大战略,加紧出台规划和政策,围绕核心技术、顶尖人才、标准规范等强化部署,力图在新一轮国际科技竞争中掌握主导权。”《新一代人工智能发展规划》将全面推动和促进智能革命的发展。

## 1.2 智能科学的兴起

人工智能研究的目标是实现人类水平的人工智能,使计算机具有人类听、说、读、写、思考、学习、适应环境变化、解决各种实际问题的能力。1977 年,曾是西蒙的研究生、斯坦福大学青年学者费根鲍姆(E Feigenbaum),在第五届国际人工智能大会上提出了知识工程(Knowledge Engineering)的概念,标志着人工智能的研究从传统的以推理为中心,进入到以知识为中心的新阶段。

知识是国家的财富,信息产业对国家的发展至关重要。1981 年 10 月日本东京召开了第五代计算机——智能计算机研讨会,东京大学元冈达教授提出了“第五代计算机的构想”。随后日本制定了研制第五代计算机的十年计划,这是一个雄心勃勃的诱人计划。1982 年夏天,日本成立了以渊一博为所长的“新一代计算机技术研究所”(ICOT)。日本通产省全力支

持该项计划，总投资预算达到 4.3 亿美元，组织富士通、NEC、日立、东芝、松下、夏普等八大著名企业配合研究所共同开发。

渊一博为所长的"新一代计算机技术研究所"苦苦奋战了将近十年。然而，"五代机"的命运是悲壮的。1992 年，因最终没能突破关键性的技术难题，无法实现自然语言人机对话、程序自动生成等目标，导致了该计划最后流产。也有人认为，"五代机"计划不能算作失败，它在前两个阶段基本上达到了预期目标。1992 年 6 月，就在"五代机"计划实施整整十年之际，ICOT 展示了它研制的五代机原型试制机，由 64 台处理器实现了并行处理，已初步具备类似人的左脑的先进功能，可以对蛋白质进行高精度分析，在基因研究中发挥了作用。

"五代机"研究失败的现实迫使人们寻找研究智能科学的新途径。智能不仅要功能仿真，而且要机理仿真；智能不仅要运用推理，自顶向下，而且要通过学习，由底向上，两者结合；脑的感知部分，包括视觉、听觉等各种感觉、运动、语言脑皮层区不仅具有输入/输出通道的功能，而且对思维活动有直接贡献。

1991 年，人工智能最权威的刊物 *Artificial Intelligence* 47 卷上发表了人工智能基础专辑，指出了人工智能研究的趋势。柯希(D Kirsh)在专辑中提出了人工智能的五个基本问题[371]：

(1) 知识与概念化是否是人工智能的核心?

(2) 认知能力能否与载体分开来研究?

(3) 认知的轨迹是否可用类自然语言来描述?

(4) 学习能力能否与认知分开来研究?

(5) 所有的认知是否有一种统一的结构?

不同学派对这些关键问题有不同的观点。各个学派从各自的优势出发，探寻人工智能走出低谷的途径。

2001 年 12 月，由美国国家科学基金会和商务部出面，组织政府部门、科研机构、大学以及工业界的专家和学者聚集华盛顿，专门研讨《提升人类能力的会聚技术》(*Converging Technologies to Improve Human Performance*)问题。以该会议提交的论文和结论为基础，2002 年 6 月，美国国家科学基金会和美国商务部共同提出了长达 468 页的《会聚技术报告》[628]。报告认为：认知科学、生物学、信息学和纳米科技等在当前为迅猛发展的领域，这四个科学及相关技术的有机结合与融合形成会聚技术。认知领域，包括认知科学与认知神经科学；生物领域，包括生物技术、生物医药及遗传工程；信息领域，包括信息技术及先进计算和通信；纳米领域，包括纳米科学和纳米技术。这些学科各自独特的研究方法与技术的融合，将加速人类对智能科学以及相关学科的研究，最终推动社会的发展。会聚技术的发展将显著提高生命质量，提升和扩展人的能力，使整个社会的创新能力和国家的生产力水平大大提高，从而增强国家的竞争力，也将对国家安全提供更强有力的保障。

20 世纪，生命科学与信息技术结合，导致生物信息学的形成和发展。21 世纪，在会聚技术的推动下，生命科学与信息技术结合，诞生了交叉学科智能科学，为智能革命指明发展的途径。2002 年，中国科学院计算技术研究所智能科学实验室创建了世界第一个智能科学网站：http://www.intsci.ac.cn/。2003 年，笔者提出智能科学研究智能的基本理论和实现技术，是由脑科学、认知科学、人工智能等学科构成的交叉学科。脑科学从分子水平、细胞水平、行为水平研究人脑智能机理，建立脑模型，揭示人脑的本质。认知科学是研究人类感知、

学习、记忆、思维、意识等人脑心智活动过程的科学。人工智能研究用人工的方法和技术,模仿、延伸和扩展人的智能,实现机器智能。三门学科共同研究,探索智能科学的新概念、新理论、新方法,必将在 21 世纪共创辉煌。

我们要向人脑学习,研究人脑智能的机制和算法成为当今研究热点。2013 年 1 月 28 日,欧盟启动了旗舰“人类大脑计划”(Human Brain Project)。2013 年 4 月 2 日,美国启动“运用先进创新型神经技术的大脑研究”(Brain Research through Advancing Innovative Neurotechnologies,BRAIN)计划。我国也在积极筹备“脑科学与类脑研究计划”。

## 1.3 脑科学

人脑是世界上最复杂的物质,它由数百种不同类型的上千亿的神经细胞所构成。理解大脑的结构与功能是 21 世纪最具挑战性的前沿科学问题;理解认知、思维、意识和语言的神经基础,是人类认识自然与自身的终极挑战[907]。现代神经科学的起点是神经解剖学和组织学对神经系统结构的认识和分析。从宏观层面,布洛卡(Paul Broca)和韦尼克(Wernicke)对大脑语言区的定位,布罗德曼(Brodmann)对脑区的组织学分割,彭菲尔德(Penfield)对大脑运动和感觉皮层对应身体各部位的图谱绘制、功能核磁共振成像对在活体进行定位时脑内依赖于电活动的血流信号等,使我们对大脑各脑区可能参与某种脑功能已有相当的理解。神经元种类图谱、介观神经联接图谱、介观神经元电活动图谱的制作将是脑科学界长期的工作。

神经系统和脑的功能从本质上是接收内外环境中的信息,加以处理、分析和存储,然后控制调节机体各部分,做出适当的反应。因此,神经系统和脑是两种活的信息处理系统。从神经元的真实生物物理模型、它们的动态交互关系以及神经网络的学习,到脑的组织和神经类型计算的量化理论等,从计算角度理解脑;研究非程序的、适应性的、大脑风格的信息处理的本质和能力,探索新型的信息处理机理和途径,从而创造脑。

计算神经科学的研究源远流长。1875 年意大利解剖学家戈尔吉(C Golgi)用染色法最先识别出单个的神经细胞。1889 年卡贾尔(R Cajal)创立神经元学说,认为整个神经系统是由结构上相对独立的神经细胞构成。在卡贾尔神经元学说的基础上,1906 年谢灵顿(C S Sherrington)提出了神经元间突触的概念[676]。1907 年拉皮克(Lapique)提出整合放电(Integrate-and-Fire)神经元模型。20 世纪 20 年代阿德廉(E D Adrian)提出神经动作电位。1943 年麦克鲁奇(W S McCulloch)和皮兹提出了的 M-P 神经网络模型[483]。1949 年赫布(D O Hebb)提出了神经网络学习的规则[311]。霍奇金(A L Hodgkin)和哈斯利(A F Huxley,)于 1952 年提出 Hodgkin-Huxley 模型[325],描述细胞的电流和电压的变化。20 世纪 50 年代罗森勃拉特(F Rosenblatt)提出了的感知机模型[632]。20 世纪 80 年代以来,神经计算研究取得了进展。霍普菲尔德(J J Hopfield)引入李雅普诺夫(Lyapunov)函数(又叫做“计算能量函数”)给出了网络稳定判据[330],可用于联想记忆和优化计算。甘利俊一(Amari)在神经网络的数学基础理论方面做了大量的研究,包括统计神经动力学、神经场的动力学理论、联想记忆,特别在信息几何方面作出了奠基性的工作[14]。

计算神经科学的研究力图体现人脑的如下基本特征:①大脑皮质是一个广泛连接的巨型复杂系统;②人脑的计算是建立在大规模并行模拟处理的基础之上的;③人脑具有很强

的“容错性”和联想能力，善于概括、类比、推广；④大脑功能受先天因素的制约，但后天因素，如经历、学习与训练等起着重要作用，这表明人脑具有很强的自组织性与自适应性。人类的很多智力活动并不是按逻辑推理方式进行的，而是由训练形成的。目前，对人脑是如何工作的了解仍然很肤浅，计算神经科学的研究还很不充分。

由瑞士洛桑联邦理工大学(EPFL)马克拉姆(Henry Markram)发起的“蓝脑计划”自 2005 年开始实施，经过十年的努力，较为完整地完成了特定脑区内皮质功能柱的计算模拟。但总体而言，要真正实现认知功能的模拟还有很大鸿沟需要跨越。2013 年，马克拉姆又构思并领导筹划欧盟人脑计划(Human Brain Project，HBP)入选欧盟的未来旗舰技术项目，获得 10 亿欧元的资金支持，成为全球范围内最重要的人类大脑研究项目。该计划的目标是用超级计算机来模拟人类大脑，用于研究人脑的工作机制和未来脑疾病的治疗，并借此推动类脑人工智能的发展。参与的科学家来自欧盟各成员国的 87 个研究机构。

美国提出“运用先进创新型神经技术的大脑研究”(Brain Research through Advancing Innovative Neurotechnologies，BRAIN)计划。美国的脑计划侧重于新型脑研究技术的研发，从而揭示脑的工作原理和脑的重大疾病发生机制，其目标是像人类基因组计划那样，不仅要引领前沿科学发展，同时带动相关高科技产业的发展。在未来 10 年将新增投入 45 亿美元。BRAIN 计划提出了 9 项优先发展的领域和目标，其中依次为：鉴定神经细胞的类型并达成共识；绘制大脑结构图谱；研发新的大规模神经网络电活动记录技术；研发一套调控神经环路电活动的工具集；建立神经元电活动与行为的联系；整合理论、模型和统计方法；解析人脑成像技术的基本机制；建立人脑数据采集的机制；脑科学知识的传播与人员培训。

2014 年日本启动的“脑智”(Brain/MIND)计划，其目标是“使用整合性神经技术制作有助于脑疾病研究的大脑图谱”(Brain Mapping by Integrated Neurotechnologies for Disease Studies，Brain/MINDS)，为期 10 年，第一年 2700 万美元，以后逐年增加。此计划聚焦在使用狨猴为动物模型，绘制从宏观到微观的脑联结图谱，并以基因操作手段，建立脑疾病的狨猴模型。

中国脑计划以理解脑认知功能的神经基础为研究主体，以脑机智能技术和脑重大疾病诊治手段研发为两翼，目标是在未来 15 年内使我国的脑认知基础研究、类脑研究和脑重大疾病研究达到国际先进水平，并在部分领域起到引领作用(图 1.1)[907]

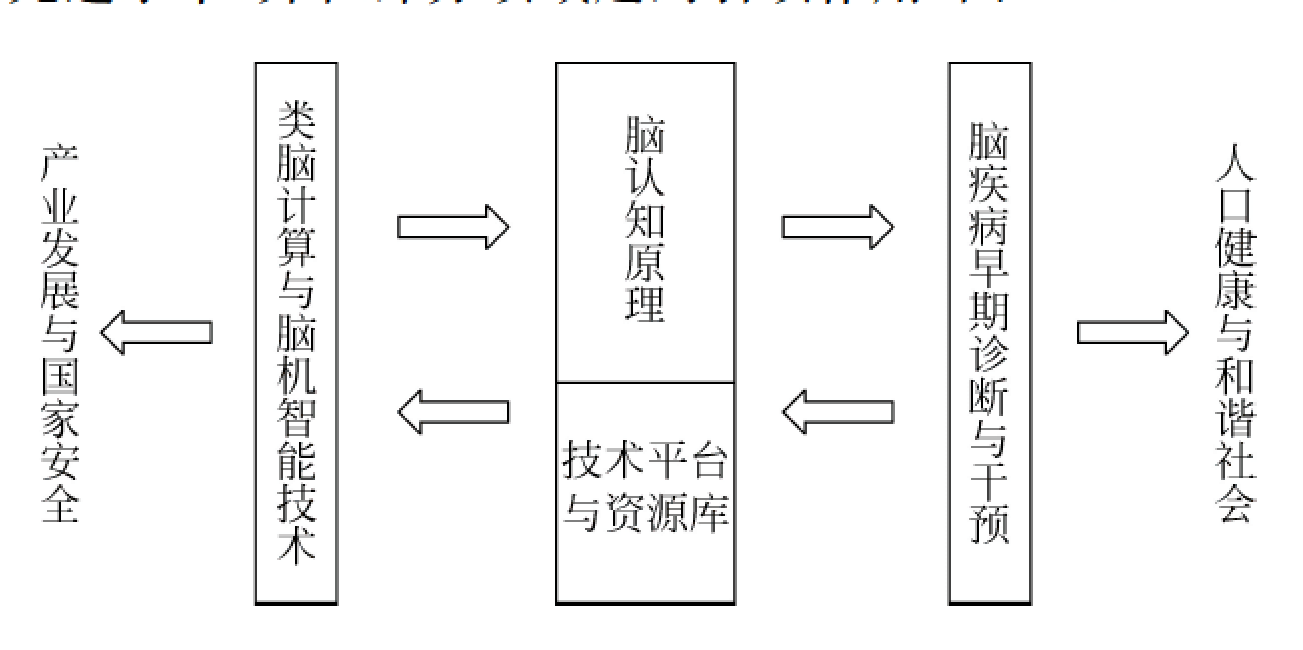



图 1.1 中国脑计划的总体格局

## 1.4 认知科学

认知是脑和神经系统产生心智的过程和活动。认知科学就是以认知过程及其规律为研究对象，探索人类的智力如何由物质产生和人脑信息处理的过程的科学。具体地说，认知科学是研究人类的认知和智力的本质和规律的前沿科学。认知科学研究的范围包括知觉、注意、记忆、动作、语言、推理、思考、意识乃至情感动机在内的各个层面的认知活动。将哲学、心理学、语言学、人类学、计算机科学和神经科学 6 大学科交叉整合，研究在认识过程中信息是如何传递的，就形成了认知科学。当前国际公认的认知科学学科结构如图 1.2 所示。认知科学的发展首先在原来的 6 个支撑学科内部产生了 6 个新的发展方向，即心智哲学、认知心理学、认知语言学(或称语言与认知)、认知人类学(或称文化、进化)、计算机科学(人工智能)和认知神经科学。这 6 个新兴学科是认知科学的 6 大学科分支。

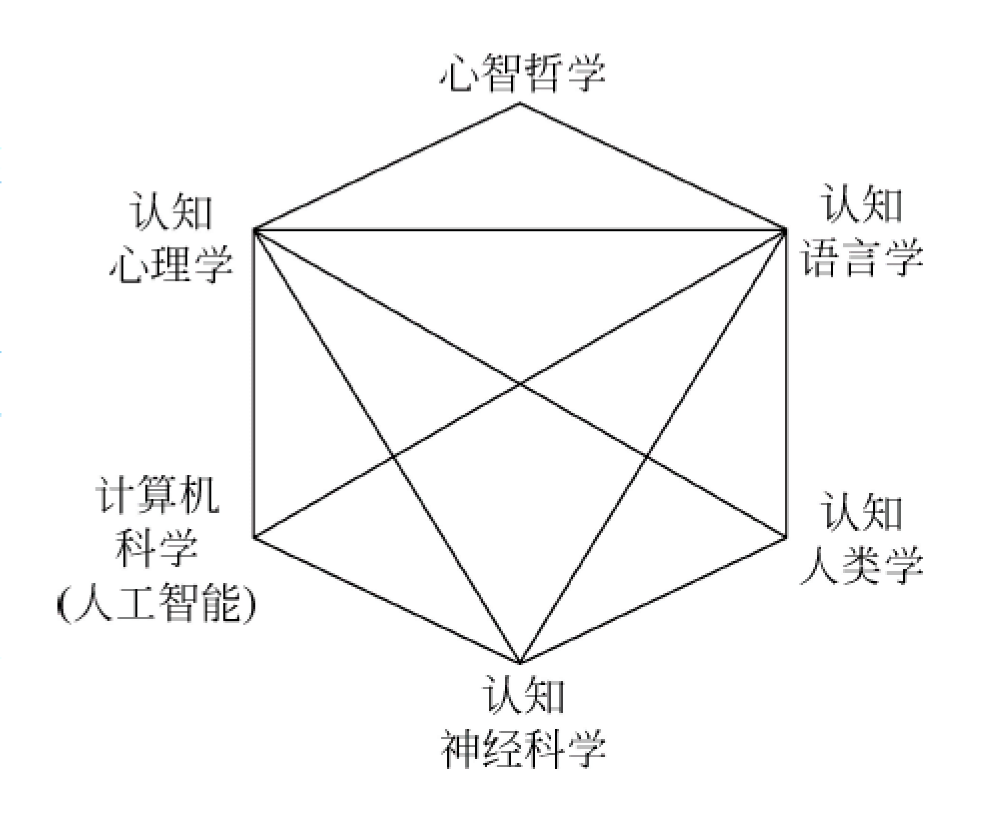



图 1.2 认知科学学科结构

最近几十年来，对复杂行为的理论主要有三个派别：新行为主义、格式塔(Gastalt)心理学派和认知心理学派。各派心理学都想更好地认识人类机体是如何活动的，它们从各个不同方面研究行为，在方法学上强调的重点不一致。新行为主义强调客观的实验方法，要求对实验严格加以控制，格式塔心理学派认为全体形态和属性并不等于各部分之和。认知心理学是用信息加工过程来解释人的复杂行为，它吸收了行为主义和格式塔心理学的有益成果。认知心理学也认为复杂的现象总要分解成最基本的部分才能进行研究。

在 20 世纪 90 年代认知科学迎来了繁荣发展的新时期。大脑成像技术的出现使得认知科学家可以观察到人们在完成各种认知任务时不同大脑区域的活动状况，认知神经科学成为认知科学当中最为活跃的领域之一。情绪、感受和意识这样一些在以往被视为“禁忌”的话题成为了认知科学研究的“热门”，认知科学的研究对象不再局限于知觉、记忆、语言、推理、学习等“狭义”的认知活动，而是力图涵盖心智的方方面面。心智不仅与大脑的结构与活动密切相关，身体也是其重要的物理基础，具身性(涉身性)成为了理解心智奥秘的关键因素之一。不仅如此，心智的边界还被延展到身体之外，物质环境和社会环境成为其不可分割的构成成分，这是延展认知和延展心智论题的基本主张。动力学系统理论则对主流认知科学的理论基础即心理表征和计算提出了强烈质疑，主张采用微分方程以及相变、吸引子、混沌等概念来刻画和理解心智的本性。从进化和适应的观点来看待人类认知能力的形成与发展，对其他动物物种认知能力的研究，也成为这一时期认知科学研究的重要课题。

认知科学的发展得到国际科技界尤其是发达国家政府的高度重视和大规模的支持。认知科学研究是“国际人类前沿科学计划”的重点。认知科学及其信息处理方面的研究被列为整个计划的三大部分之一(其余两部分是“物质和能量的转换”“支撑技术”)；“知觉和认知”“运动和行为”“记忆和学习”和“语言和思考”被列为人类前沿科学的 12 大焦点问题中的 4 个。近年来，美国和欧盟分别推出“脑的十年”计划和“欧盟脑的十年计划”。日本则推出

雄心勃勃的“脑科学时代”计划，总预算高达 200 亿美元。在“脑科学时代”计划中，脑的认知功能及其信息处理的研究是重中之重。包括知觉、注意、记忆、动作、语言、推理和思考、意识乃至情感动机在内的各个层次和各个方面的人类认知和智力活动都被列入研究的重点；将认知科学和信息科学相结合来研究新型计算机和智能系统也被列为该计划的三个方面之一。

图灵奖获得者纽厄尔以认知心理学为核心，探索认知体系结构。至今在认知心理学与人工智能领域广泛应用的认知模型 SOAR 与 ACT-R 都是在纽厄尔直接领导下或受其启发而发展起来的，并以此为基石实现了对人类各种认知功能的建模。马尔（David Marr）不但是计算机视觉的开拓者，还奠定了神经元群之间存储、处理、传递信息的计算基础，特别是对学习与记忆、视觉相关环路的神经计算建模作出了重要贡献。

认知科学和哲学家萨伽德（Paul Thagard）在《心智：认知科学导论》译本所写的前言中[754]，指出当今认知科学发展的四个新趋势。

(1) 认知神经科学的中心地位进一步得以巩固和加强，对于大脑和神经系统的更为全面和系统的研究对整个认知科学而言具有基础性的作用，甚至对一些传统哲学问题（如心身问题、自由意志和人生意义等）也具有重要的意义。

(2) 基于贝叶斯概率理论的统计模型变得日益显要，被运用于处理认知心理学当中的许多重要现象，并且在人工智能和机器人学中得到广泛应用。

(3) 具身性成为认知科学的基础性概念，心、脑、身体与物质环境和社会环境的相互作用对于理解心智的本性至关重要。

(4) 有关认知的社会的、文化的和历史的维度得到更多的重视。

## 1.5　人工智能

人工智能是通过人工的方法和技术，让机器像人一样认知、思考和学习，模仿、延伸和扩展人的智能，实现机器智能。人工智能自 1956 年诞生以来，历经艰辛与坎坷，取得了举世瞩目的成就。人工智能的发展历史可大致分为孕育期、形成期、基于知识的系统、神经网络的复兴、机器学习大发展这几个阶段。

### 1. 人工智能的孕育期(1956 年以前)

人工智能的孕育期可以认为是 1956 年以前的时期。这一时期的主要成就是数理逻辑、自动机理论、控制论、信息论、神经计算、电子计算机等学科的建立和发展，为人工智能的诞生，准备了理论和物质的基础。这一时期的主要贡献有：

(1) 1936 年图灵创立了理想计算机模型的自动机理论，提出了以离散量的递归函数作为智能描述的数学基础，给出了基于行为主义的测试机器是否具有智能的标准，即图灵测试。

(2) 1943 年，心理学家麦克洛奇和数理逻辑学家皮兹在《数学生物物理公报》上发表了关于神经网络的数学模型。这个模型，现在一般称为 M-P 神经网络模型。他们总结了神经元的一些基本生理特性，提出神经元形式化的数学描述和网络的结构方法，从此开创了神经计算的时代。

(3) 1945 年冯·诺依曼提出的存储程序概念,1946 年研制成功的第一台电子计算机 ENIAC,为人工智能的诞生奠定了物质基础。

(4) 1948 年香农发表了《通讯的数学理论》,标志着一门新学科——信息论的诞生。他认为人的心理活动可以用信息的形式来进行研究,并提出了描述心理活动的数学模型。

(5) 1948 年维纳创立了控制论。它是一门研究和模拟自动控制的生物和人工系统的学科,标志着人们根据动物心理和行为科学进行计算机模拟研究和分析的基础已经形成。

### 2. 人工智能的形成期(1956—1968)

人工智能的形成期大约从 1956 年开始到 1968 年。1956 年在美国的达特茅斯大学召开了为期两个月的学术研讨会,提出了"人工智能"这一术语,标志着这门学科的正式诞生。在定理机器证明、问题求解、LISP 语言、模式识别等关键领域取得重要进展。

(1) 1956 年,纽厄尔和西蒙的"逻辑理论家"程序,该程序模拟了人们用数理逻辑证明定理时的思维规律。这一工作受到了人们高度的评价,被认为是计算机模拟人的高级思维活动的一个重大成果,是人工智能的真正开端。

(2) 1956 年,塞缪尔研制了跳棋程序,该程序具有学习功能,能够从棋谱中学习,也能在实践中总结经验,提高棋艺。它在 1959 年打败了塞缪尔本人,又在 1962 年打败了美国一个州的跳棋冠军。这是模拟人类学习过程的一次卓有成效的探索,是人工智能的一个重大突破。

(3) 1958 年,麦卡锡提出表处理语言 LISP,不仅可以处理数据,而且可以方便地处理符号,成为人工智能程序设计语言的重要里程碑。目前 LISP 语言仍然是人工智能系统重要的程序设计语言和开发工具。

(4) 1960 年,纽厄尔、肖(Shaw)和西蒙等人研制了通用问题求解程序 GPS,是对人们求解问题时的思维活动的总结。首次提出了启发式搜索的概念。

(5) 1965 年鲁宾逊(J A Robinson)提出归结法,被认为是一个重大的突破,也为定理证明的研究带来了又一次高潮。

(6) 1968 年,斯坦福大学费根鲍姆等人研制成功了化学分析专家系统 DENDRAL,被认为是专家系统的萌芽,是人工智能研究从一般思维探讨到专门知识应用的一次成功尝试。

此外还有许多其他的成就,如 1956 年乔姆斯基(N Chomsky)提出的文法体系等。

### 3. 基于知识的系统(1968—1992)

1965 年,斯坦福大学的费根鲍姆和化学家勒德贝格(J Lederberg)合作研制 DENDRAL 系统。1972—1976 年,布坎南(Bruce Buchanan)和费根鲍姆开发成功医疗专家系统 MYCIN。此后,许多著名的专家系统相继研发成功,其中较具代表性的有探矿专家系统 PROSPECTOR、青光眼诊断治疗专家系统 CASNET、钻井数据分析专家系统 ELAS 等。20 世纪 80 年代,专家系统的开发趋于商品化,创造了巨大的经济效益。

1977 年,费根鲍姆在第五届国际人工智能联合会议(IJCAI'77)上提出知识工程的新概念。他认为,"知识工程是人工智能的原理和方法,对那些需要专家知识才能解决的应用难题提供求解的手段。恰当运用专家知识的获取、表达和推理过程的构成与解释,是设计基于知识的系统的重要技术问题。"知识工程是一门以知识为研究对象的学科,它将具体智能系

统研究中那些共同的基本问题抽出来，作为知识工程的核心内容，使之成为指导具体研制各类智能系统的一般方法和基本工具。

知识工程的兴起，确立了知识处理在人工智能学科中的核心地位，使人工智能摆脱了纯学术研究的困境，使人工智能的研究从理论转向应用，从基于推理的模型转向知识的模型，使人工智能的研究走向了实用。

进入 20 世纪 80 年代，日本在 1981 年宣布了第五代计算机的研制计划。其研制的计算机主要特征是具有智能接口、知识库管理、自动解决问题的能力，并在其他方面具有人的智能行为。由于这一计划的提出，形成了一股热潮，促使世界上重要的国家都开始制定对新一代智能计算机的开发和研制计划，使人工智能进入了一个基于知识的兴旺时期。

#### 4. 神经网路的复兴(1982—)

1958 年，罗森勃拉特提出感知器作为模式识别问题的新方法。在 20 世纪 60 年代，感知器神经网络好像可以做任何事。明斯基和佩珀特(Papert)在 1969 年合写的《感知器》(*Perceptrons*)一书中，利用数学理论证明单层感知器所能计算的本质局限[508]。这使全世界中止了神经网络的研究计划，人们常称神经网络研究进入“黑暗时期”。

1982 年，美国加州工学院物理学家霍普菲尔德使用统计力学的方法来分析网路的存储和优化特性，提出了离散的神经网络模型，从而有力地推动了神经网络的研究。1984 年霍普菲尔德又提出了连续神经网络模型。

20 世纪 80 年代神经网路复兴的真正推动力是反向传播算法的重新研究。该算法最早由哈佛大学的韦勃斯(P J Werbos)于 1974 年 8 月提出。1986 年，鲁梅尔哈特(D E Rumelhart)和麦克莱伦德(J L McClelland)等人提出并行分布处理(Parallel Distributed Processing，PDP)的理论，致力于认知的微观结构的探索，其中多层网络的误差传播学习法，即反向传播算法广为流传，引起人们极大的兴趣。世界上许多国家掀起了神经网络研究的热潮。从 1985 年开始，专门讨论神经网络的学术会议规模逐步扩大。1987 年在美国召开了第一届神经网络国际会议，并发起成立国际神经网络学会(INNS)。

#### 5. 机器学习大发展(20 世纪 90 年代中期—)

数据是载体，智能是目标，而机器学习是从数据通往智能的技术途径。Boosting、支持向量机(SVM)、集成学习和稀疏学习是机器学习界也是统计界在近十年或者是近二十年来最为活跃的方向，这些成果是统计界和计算机科学界共同努力成就的。数学家瓦普尼克(V N Vapnik)于 1995 年提出支持向量机，成为分类算法的一个基准模型[774]。核主成分分析(Kernel Principal Component Analysis，KPCA)是由计算机学家提出的一个非线性降维方法，其实它等价于经典多维尺度分析(Multi-Dimensional Scaling，MDS)。2011 年的图灵奖授予了加州大学洛杉矶分校教授珀尔(Judea Pearl)，他主要的研究领域是概率图模型和因果推理[551]。这是机器学习的基础问题。

2006 年，多伦多大学教授辛顿(G E Hinton)在前向神经网络的基础上，提出了深度学习[320]。深度学习在 AlphaGo、无人驾驶汽车、人工智能助理、语音识别、自然语言理解等方面取得很好进展，对工业界产生了巨大影响。

从 20 世纪 60 年代以来，冯·诺依曼体系结构是计算机体系结构的主流，计算机运算速

度得到极大提高。"神威·太湖之光"超级计算机运行速度超过十亿亿次/秒,其计算能力相当于 200 多万台普通电脑的运算,运算 1 分钟相当于 72 亿地球人同时用计算器不间断计算 32 年。但是计算机的智能水平仍很低下。在经典的计算机中,将数据处理的地方与数据存储的地方分开,存储器和处理器通过总线连接。大多数现代处理器在扩大寄存器的同时使用缓存,以在靠近计算点的地方提供临时的存储。高速缓存的架构对这种模拟脑的计算挑战不起作用。即使是相对简单的脑也是由几十亿个突触连结的几千万个神经元组成,因此,要模拟这样庞大的相互联系的脑需耗费与计算机主存储器容量一样大的高速缓存,这会导致机器立即无法使用。人脑的前额叶工作记忆区兼具存储和运算的功能。

现有的人工智能技术缺乏通用性。语音识别、图像处理、自然语言处理、专家系统等采用不同的模型和不同的学习数据,两种不同的任务无法采用同一套系统进行求解,不同任务之间知识也无法共享。而人脑却采用同一个信息处理系统进行自动感知、问题分析与求解、决策控制等。

目前基于深度学习方法已经在人工智能领域取得了很大的成功,然而其代价是高能耗。例如 AlphaGo 运行时,共用了 1202 个 CPU 核和 176 个 GPU 核,采用了 40 个搜索线程,能耗大。而人类大脑的功耗仅在 20W 左右。

我们要向人脑学习,研究人脑信息处理的方法和算法。人脑信息处理过程不再仅凭猜测,而通过多学科交叉和实验研究获得人脑的工作机制。因此,受脑信息处理机制启发,借鉴脑神经机制和认知行为机制发展智能科学,已成为近年来人工智能与信息科学领域的研究热点。智能科学方兴未艾,引领人工智能和智能技术蓬勃发展。

## 1.6 智能科学的研究内容

智能科学的研究内容包括计算神经理论、认知计算、知识工程、自然语言处理、智能机器人等。

### 1. 计算神经理论

脑是一个由神经元构成的网络,神经元与神经元之间的相互联系依赖于突触,这些彼此联系的神经元构成一定的神经网路来发挥大脑的功能。这些相互作用对神经环路功能的稳态平衡、复杂性以及信息加工处理中发挥着关键作用。与此同时,神经元膜上的受体和离子通道对于控制神经元的兴奋性、调节突触功能以及神经元内各种递质和离子的动态平衡至关重要。认识大脑的神经网络结构及其形成复杂认知功能的机制是认识、开发和利用脑的基础。

计算神经理论从分子水平、细胞水平、行为水平研究知识和外界事物在脑内如何表达、编码、加工和解码,揭示人脑智能机理,建立脑模型。需要研究的问题有神经网络是如何形成的?中枢神经系统是如何构建的?在神经网络形成的过程中,神经细胞的分化、神经元的迁移、突触的可塑性、神经元活动的与神经递质和离子通道、神经回路形成以及信息的整合等。这些问题的研究将对智能机理提供有力的脑科学基础。

### 2. 认知计算

认知计算从微观、介观、宏观等不同尺度上研究人脑如何实现感知、学习、记忆、思维、情感、意识等心智活动。感知是人们对客观事物的感觉和知觉过程。感觉是人脑对直接作用于感觉器官的客观事物的个别属性的反映。知觉是人脑对直接作用于感觉器官的客观事物的整体的反映。知觉信息的表达、整体性、知觉的组织与整合属于知觉研究的基本问题，是研究其他各个层次认知的基础。迄今为止已建立四种知觉理论：构造论者的探讨对于学习和记忆的因素赋予较大的影响，认为所有感知都受到人们的经验和期望的影响；吉布森(J J Gibson)的生态学着重探讨在刺激模式中所固有的全部环境的信息，认为知觉是直接的，没有任何推理步骤、中介变量或联想。格式塔理论偏重强调知觉组织的先天论的因素，提出整体大于局部之和；动作理论集中于探讨知觉者在他的环境中做动作探测所产生的反馈作用。模式识别是人类的一项基本智能。模式识别研究主要集中在两方面，一是研究生物体(包括人)是如何感知对象的；二是在给定的任务下，如何用计算机实现模式识别的理论和方法。

学习是基本的认知活动，是经验与知识的积累过程，也是对外部事物前后关联地把握和理解以便改善系统行为的性能的过程。学习理论是指有关学习的实质、学习的过程、学习的规律以及制约学习的各种条件的理论探讨和解释。在探讨学习理论的过程中，由于各自的哲学基础、理论背景、研究手段的不同，自然形成了各种不同的理论观点，并形成了各种不同的理论派别，主要包括行为学派、认知学派和人本主义学派。

学习的神经生物学基础是神经细胞之间的联系结构突触的可塑性变化，已成为当代神经科学中一个十分活跃的研究领域。突触可塑性条件即在突触前纤维与相联的突后细胞同时兴奋时，突触的连接加强。1949 年，加拿大心理学家赫布提出了 Hebb 学习规则[311]。他设想在学习过程中有关的突触发生变化，导致突触连接的增强和传递效能的提高。Hebb 学习规则成为连接学习的基础。

记忆就是对过去的经验或是经历，在脑内产生准确的内部表征，并且能够正确、高效地提取和利用它们。记忆涉及信息的获得、储存和提取等多个过程，这也就决定了记忆需要不同的脑区协同作用。在最初的记忆形成阶段，需要脑整合多个分散的特征或组合多个知识组块以形成统一的表征。从空间上讲，不同特征的记忆可能储存于不同的脑区和神经元群；而在时间上，记忆的储存又分为工作记忆、短时记忆和长时记忆。

研究工作记忆的结构与功能，对认识人的智能的本质具有重大意义。1974 年，巴德利(A D Baddeley)和希契(G J Hitch)在模拟短时记忆障碍的实验基础上提出了工作记忆的三系统概念，用“工作记忆”代替了原来“短时记忆”的概念。巴德利认为工作记忆指的是一种系统，它为复杂的任务比如言语理解、学习和推理等提供临时的储存空间和加工时所必需的信息。工作记忆系统能同时储存和加工信息，这和短时记忆概念仅强调储存功能是不同的[51]。人们发现工作记忆与语言理解能力、注意及推理等联系紧密，工作记忆蕴藏智能的玄机。

思维是具有意识的人脑对于客观现实的本质属性、内部规律性的自觉的、间接的、概括的反映，以内隐或外显的语言或动作表现出来。思维是由复杂的脑机制所赋予的，对客观的关系、联系进行着多层加工，揭露事物内在的、本质的特征，是认知的高级形式。人类思维的

形态主要有抽象(逻辑)思维、形象(直感)思维、感知思维和灵感(顿悟)思维。思维的研究对理解人类的认知和智力的本质,对人工智能的发展将具有重要的科学意义和应用价值。通过研究不同层次的思维模型,研究思维的规律和方法,为新型智能信息处理系统提供原理和模型。

人工智能的奠基人之一明斯基认为情感是人类一种特殊的思维方式,指出没有情感的机器怎么能是智能的? 因此,让计算机具有情感,也就是让计算机更加智能。情感计算领域的创始人皮卡德(R W Picard)把“情感计算”定义为“与情感有关、由情感引发或者能够影响情感的因素的计算”[575]。情感计算是建立和谐人机环境的基础之一,其目的是赋予计算机识别、理解、表达和适应人情感的能力,提高人机交互的质量和效率。目前,情感计算研究受到广泛关注。MIT 情感计算研究小组开发可穿戴计算机,识别真实情境中的人类情感,研究人机交互中的情感反馈机制,研制能够用肢体语言表达情感的机器人。瑞士政府成立了情感科学中心,心理学、神经科学、哲学、历史学、经济、社会、法律等多学科合作开展情感计算的研究与应用。日本文部省曾支持“情感信息的信息学、心理学研究”重点基金项目。日内瓦大学成立了情绪研究实验室,布鲁塞尔大学则建立了情绪机器人研究小组。英国伯明翰大学开展了“认知与情感的研究”。欧盟也把情感计算列入研究计划。情感计算的研究除了为人工智能的发展提供一条新的途径之外,同时对于理解人类的情绪,乃至人类的思维都有着重要的价值,关于情感本身及情感与其他认知过程间相互作用的研究成为智能科学的研究热点。

意识是生物体对外部世界和自身心理、生理活动等客观事物的觉知。意识是智能科学研究的核心问题。为了揭示意识的科学规律,建构意识的脑模型,不仅需要研究有意识的认知过程,而且需要研究无意识的认知过程,即脑的自动信息加工过程,以及两种过程在脑内的相互转化过程。同时,自我意识与情境意识也是需要重视的问题。自我意识(Self Consciousness)是个体对自己存在的觉察,是自我知觉的组织系统和个人看待自身的方式,包括自我认知、自我体验、自我控制三种心理成分。情境感知(Situation Awareness)是个体对不断变化的外部环境的内部表征。在复杂、动态变化的社会信息环境中,情境感知是影响人们决策和绩效的关键因素。意识的认知原理,意识的神经生物学基础以及意识与无意识的信息加工等是需要重点研究的问题。

心智(Mind)是人类全部精神活动,包括情感、意志、感觉、知觉、表象、学习、记忆、思维、直觉等,用现代科学方法来研究人类非理性心理与理性认知融合运作的形式、过程及规律。建立心智模型的技术常称为心智建模,目的是为了从某些方面探索和研究人的思维机制,特别是人的信息处理机制,同时也为设计相应的人工智能系统提供新的体系结构和技术方法。心智问题是一个非常复杂的非线性问题,我们必须借助现代科学的方法来研究心智世界。

### 3. 知识工程

知识工程研究知识的表示、获取、推理、决策和应用,包括大数据、机器学习、数据挖掘和知识发现、不确定性推理、知识图谱、机器定理证明、专家系统、机器博弈、数字图书馆等。

大数据(Big Data),指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合,是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。在迈尔-舍恩伯格(Viktor Mayer-Schönberger)及库克耶

(K Cukier)编写的《大数据时代》[471]中,大数据指不用随机分析法(抽样调查)这样的捷径,而采用所有数据进行分析处理。IBM 提出大数据的 5V 特点：Volume(大量)、Velocity(高速)、Variety(多样)、Value(低价值密度)、Veracity(真实性)。

机器学习研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习方法有归纳学习、类比学习、分析学习、强化学习、遗传算法、连接学习和深度学习等。

### 4. 自然语言处理

人类进化过程中,语言的使用使大脑两半球功能分化。语言半球的出现使人类明显有别于其他灵长类。一些研究表明,人脑左半球同串行的、时序的、逻辑分析的信息处理有关,而右半脑同并行的、形象的、非时序的信息处理有关。

语言是以语音为外壳、以词汇为材料、以语法为规则而构成的体系。语言通常分为口语和文字两类。口语的表现形式为声音,文字的表现形式为形象。口语远较文字古老,个人学习语言也是先学口语,后学文字。语言是最复杂、最系统而且应用又最广的符号系统。语言符号不仅表示具体的事物、状态或动作,而且也表示抽象的概念。汉语以其独特的词法和句法体系、文字系统和语音声调系统而显著区别于印欧语言,具有音、形、义紧密结合的独特风格。从神经、认知和计算三个层次上研究汉语,给予我们开启智能之门极好的机遇。

自然语言理解实现人与计算机之间用自然语言进行有效通信的各种理论和方法,要研究自然语言的语境、语义、语用和语构,包括语音和文字的计算机输入,大型词库、语料和文本的智能检索,机器语音的生成、合成和识别,不同语言之间的机器翻译和同传等。

### 5. 智能机器人

智能机器人拥有相当发达的“人工大脑”,可以按目的安排动作,还具有传感器和效应器。智能机器人研究可以分为基础前沿技术、共性技术、关键技术与装备、示范应用四个层次。其中基础前沿技术主要涉及机器人新型机构设计、智能发育理论与技术,以及互助协作型、人体行为增强型等新一代机器人验证平台研究等。共性技术主要包括核心零部件、机器人专用传感器、机器人软件、测试/安全与可靠性等关键共性技术研发。关键技术与装备主要包括工业机器人、服务机器人、特殊环境服役机器人和医疗/康复机器人的关键技术与系统集成平台研发。示范应用面向工业机器人、医疗/康复机器人等领域的示范应用等。20 世纪末,计算机文化已深入人心。21 世纪,机器人文化将对社会生产力的发展,对人类生活、工作、思维的方式以及社会发展产生无可估量的影响。

## 1.7　展望

智能科学是生命科学技术的精华、信息科学技术的核心,现代科学技术的前沿和制高点,涉及自然科学的深层奥秘,触及哲学的基本命题。因此,一旦取得突破,将对国民经济、社会进步、国家安全产生特别深刻、特别巨大的影响。目前,智能科学正处在方法论的转变期、理论创新的高潮期和大规模应用的开创期,充满原创性机遇。

智能科学的目标旨在探索智能的本质,建立智能科学和新型智能系统的计算理论,解决

对智能科学和信息科学具有重大意义的基础理论和智能系统实现的关键技术问题,将在类脑智能机、智能机器人、脑机融合、智能系统等方面得到广泛的应用。

人类文明发展到现在,共发生了五次科技革命。历次科技革命的影响,可以从以下三方面来评价。

(1) 对人类的生活方式和思维方式影响:第一次科技革命主要包括新物理学诞生,近代科学的全面发展;第二次科技革命主要是蒸汽机、纺织机等的出现,机器代替人力;第三次科技革命主要是发电机、内燃机、电信技术的出现,同时我们的生存空间获得极大扩展;第四次科技革命是现代科学的开端,主要是认知空间的极大扩展;第五次科技革命是信息革命,社会交流方式和信息获取方式的极大发展。

(2) 重大的理论突破:第一次科技革命主要产生了哥白尼学说、伽利略学说以及牛顿力学;第二次科技革命是热力学卡诺理论、能量守恒定律等的建立;第三次科技革命主要是电磁波理论的建立;第四次科技革命主要是进化论、相对论、量子论、DNA双螺旋结构理论的建立;第五次科技革命是信息革命,产生了冯·诺依曼理论、图灵理论。

(3) 对经济和社会的影响:第一次科技革命中科学的启蒙为未来的机械革命等奠定了理论基础;第二次科技革命开始了以工厂大生产方式为特征的工业革命;第三次科技革命拓展了新兴市场,开拓了现代化的工业时代;第四次科技革命推动了20世纪绝大部分的科技文明;第五次科技革命促进了经济全球化,大数据时代到来。

当今世界科技正处于新一轮革命性变革的拂晓。第六次科技革命将是智能革命,用机器取代或增强人类的智力劳动。在第六次科技革命中,智能技术将起主导作用,智能科学将引领其发展。

2017年7月8日,国务院发布了《新一代人工智能发展规划》的通知。通知指出人工智能成为国际竞争的新焦点,是经济发展的新引擎。它将深刻改变人类生产生活方式和思维模式,实现社会生产力的整体提升。我们要牢牢把握人工智能和智能科学发展的重大历史机遇,引领世界人工智能和智能科学发展新潮流,带动国家竞争力整体提升和跨越式发展。

# 第 2 章<<
# 神经生理基础

人脑是世界上最复杂的物质，它是人类智能与高级精神活动的生理基础。脑是认识世界的器官，要研究人类的认知过程和智能机理，就必须了解这个高度复杂而有序的物质的生理机制。脑科学和神经科学从分子水平、细胞水平、行为水平研究自然智能机理，建立脑模型，揭示人脑的本质，极大地促进智能科学的发展。神经生理及神经解剖是神经科学的两大基石。神经解剖学介绍神经系统的构造，神经生理学则介绍神经系统的功能。本章主要介绍智能科学的神经生理基础。

## 2.1 脑系统

人脑由前脑、中脑、后脑所组成（图 2.1）。脑的各部分具有不同的功能，并有层次上的差别。脑的任何部分都与大脑皮质有联系，通过这种联系，把来自各处的信息汇集在大脑皮质进行加工、处理。前脑包括大脑半球和间脑。

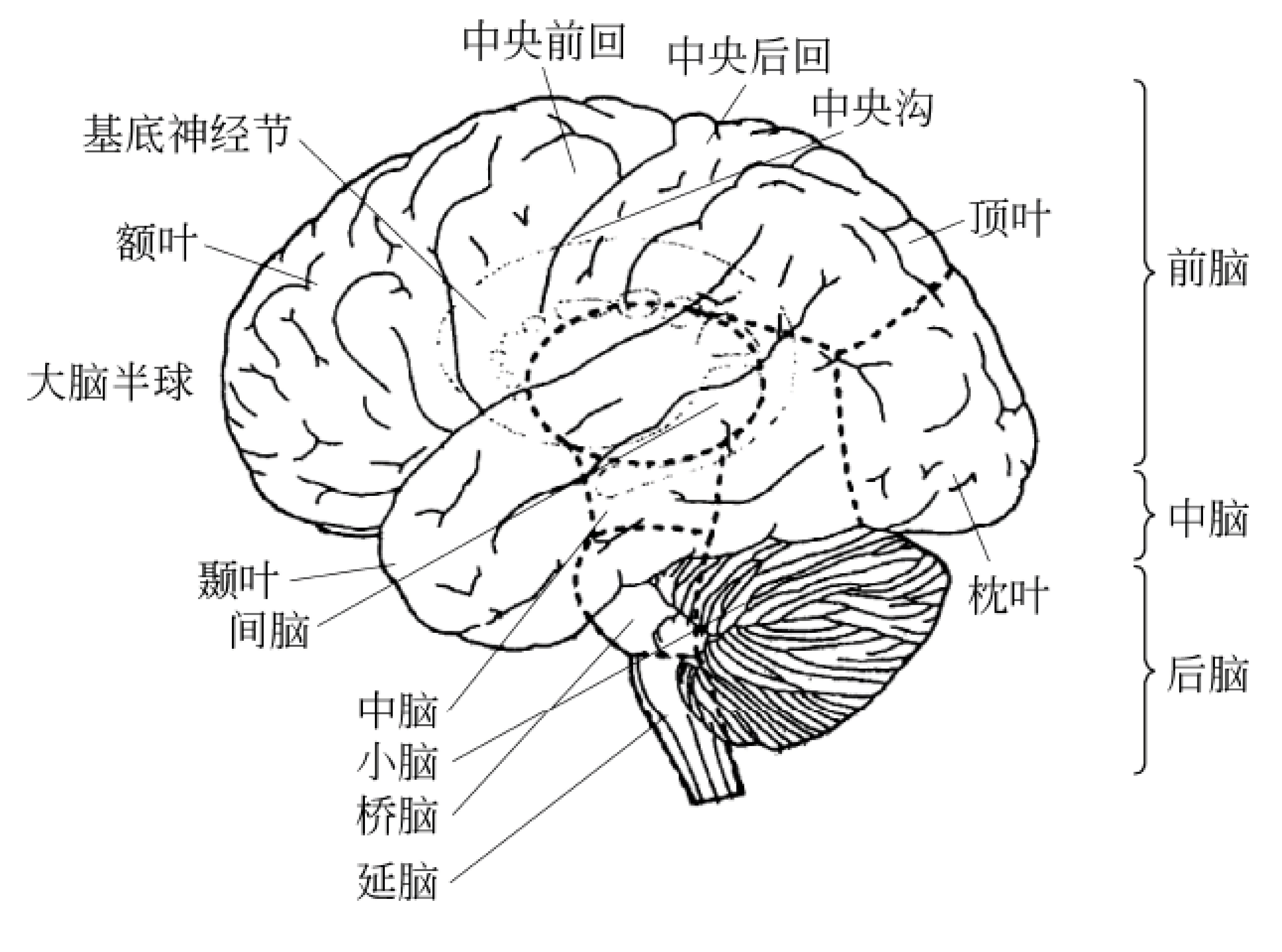



图 2.1　脑系统

（1）大脑（Cerebrum）　由左右两个大脑半球构成。其间留有一纵裂，裂的底部由被称为胼胝体的横行纤维连接。两半球内均有间隙，左右对称，称侧脑室。半球表面层为灰质，称为大脑皮质，表面有许多沟和回，增加了皮层的表面面积；内层为髓质，髓质内藏有灰质核团，为基底神经节、海马和杏仁核。大脑皮质分为额叶、颞叶、顶叶和枕叶。

（2）间脑（Diencephalon）　是围成第三脑室的脑区。上壁很薄，由第三脑室脉络丛构成。两侧壁上部的灰质团称丘脑。丘脑背面覆盖一薄层纤维，称带状层。在丘脑内部有与

此带状层相连的 Y 形白质板称内髓板，将丘脑分为前、内和外侧三大核团。上丘脑位于第三脑室顶部周围，下丘脑包括第三脑室侧壁下部的一些核团，位于丘脑的前下方。后丘脑是丘脑向后的延伸部，由内（与听觉有关）与外（与视觉有关）膝状体构成，还有底丘脑为间脑与中脑尾侧的移行地带。丘脑编码和转输传向大脑皮质的信息；下丘脑协调植物性、内分泌和内脏功能。

（3）中脑（Mesencephalon） 由大脑脚和四叠体构成，协调感觉与运动功能。

（4）后脑（Metencephalon） 由桥脑、小脑、延脑构成。小脑由蚓部和两侧的小脑半球构成；协调运动功能。桥脑宛如将两侧小脑半球连起来的桥，主要传输从大脑半球向小脑的信息。延脑介于桥脑与脊髓之间，是控制心跳、呼吸和消化等植物的神经中枢。桥脑与延脑的背侧面共同形成第四脑室底，呈菱形窝，窝顶为小脑所覆盖，即由三者共同围成第四脑室。此脑室上接中脑水管与第三脑室相通，下与脊髓中央管相通。

人脑是一个结构复杂且功能齐全的系统。统观大脑两半球的全局，可以把它划分为几个具有不同机能的区域，枕叶位于大脑半球的后部，是视区，对视觉刺激进行分析、综合；顶叶在枕叶之前，顶叶前部对触觉刺激以及在肌肉和关节器官中发生的刺激进行分析和综合；颞叶在枕叶下前方，颞叶的上部对来自听觉器官的刺激进行分析、综合；额叶位于大脑半球的前部，面积最大，额叶的后部报道关于身体的运动和它在空间位置的信号。分别研究大脑两半球就会发现，大脑两半球具有两套信息加工系统，它们的神经网络分别以不同方式来反映世界。对大多数人而言，左半球在语言、逻辑思维、数学计算和分析能力方面起主导作用；右半球则善于解决空间问题，主管音乐、美术、直观的创造性的综合性活动。正常时，这两种方式互相穿插、转化，形成整个人脑对客观世界的统一而完善的认识。

现代神经生理学家认为，脑的高级功能的出现与神经网络的活动有着密切的关系。例如，美国著名神经生理学家、诺贝尔奖获得者斯佩里（R Sperry）就十分明确地说，他认为主观的意识和思维是脑过程的一个组成部分，取决于神经网路及其有关的生理特性，是脑的高层次活动的结果。法国的神经生理学家尚格也说，行为、思维和情感等来源于大脑中产生的物理和化学现象，是相应神经元组合的结果。

真正的神经科学起始于 19 世纪末。1875 年意大利解剖学家戈尔吉用染色法最先识别出单个的神经细胞。1889 年卡贾尔创立神经元学说，认为整个神经系统是由结构上相对独立的神经细胞构成。近几十年来，神经科学和脑功能研究的发展极为迅速，并取得进展。据估计，整个人脑神经元的数量约为 $10^{11}$（千亿）。每个神经元由两部分构成：神经细胞体及其突起（树突和轴突）。细胞体的直径从 5～100$\mu$m 不等。各个神经细胞发出突起的数目、长短和分支也各不相同。长的突起可达 1m 以上，短的突起则不到其千分之一。神经元之间通过突触互相连接。突触的数量是惊人的。据测定，在大脑皮质的一个神经元上，突触的数目可达 3 万以上。整个脑内突触的数目约在 $10^{14}$～$10^{15}$（百万亿～千万亿）。突触联系的方式是多种多样的，常见的是一个神经元的纤维末梢与另一个神经元的胞体或树突形成突触联系。但也有轴突与轴突、胞体与胞体以及其他方式的突触联系。不同方式的突触连接，其生理作用是不同的。

神经元之间的组合形式也是多种多样的。一个神经元可以通过纤维分支与许多神经元建立突触联系，使得一个神经元的信息可以直接传递给许多神经元。不同部位、不同区域的神经元的纤维末梢也可汇聚到一个神经元上，使得不同来源的信息集中到一起。此外，还有

环形组合、链形组合等。因此,神经元之间的联系是十分错综复杂的。

神经网络的复杂多样,不仅在于神经元和突触的数量大、组合方式复杂和联系广泛,还在于突触传递的机制复杂。现在已经发现和阐明的突触传递机制有:突触后兴奋,突触后抑制,突触前抑制,突触前兴奋,以及“远程”抑制等。在突触传递机制中,释放神经递质是实现突触传递机能的中心环节,而不同的神经递质有着不同的作用、性质和特点。

人脑是漫长的生物演化过程的产物。动物界在进化历程中花了大约 10 亿年的时间。单细胞生物无所谓神经系统。到了扁虫类,神经细胞开始集中在头部而形成神经节。动物大脑的最初分化和嗅觉有关,两栖类和鱼类以下的动物,只有和嗅觉密切相连的嗅叶。从脊椎动物开始,有了中枢神经系统,鱼类的脑已有了端脑、间脑、中脑、后脑和延脑这 5 个部分。在爬行类动物中,大脑新皮质开始出现,真正的大脑即新皮质见于哺乳动物。灵长类的大脑皮质得到了充分的发展,掌管了对机体各种机能的全面而又精细的调节。在这个进化过程的终端,形成了极其复杂的神经网络,构成了巨系统的思维器官——人脑。

人脑的研究已成为科学研究的前沿。有的专家估计,继诺贝尔生理学——医学奖获得者沃森(J D Watson)和克里克(F Crick)于 20 世纪 50 年代提出 DNA 分子双螺旋结构,成功地解释了遗传学问题,在生物学中掀起分子生物学研究的浪潮以后,脑科学将是下一个浪潮。西方许多从事生物学、物理学研究的一流科学家在得到诺贝尔奖后纷纷转入脑科学研究。

## 2.2　神经组织

神经系统的主要细胞组成是神经细胞和神经胶质细胞。神经系统表现出来的一切兴奋、传导和整合等机能特性都是神经细胞的机能。胶质细胞占脑容积一半以上,数量大大超过了神经细胞,但在机能上只起辅助作用。

### 2.2.1　神经元的基本组成

神经细胞是构成神经系统最基本的单位,故通称为神经元。一般包括神经细胞体(Soma)、轴突(Axon)和树突(Dendrites)三部分。神经元的一般结构如图 2.2 所示。

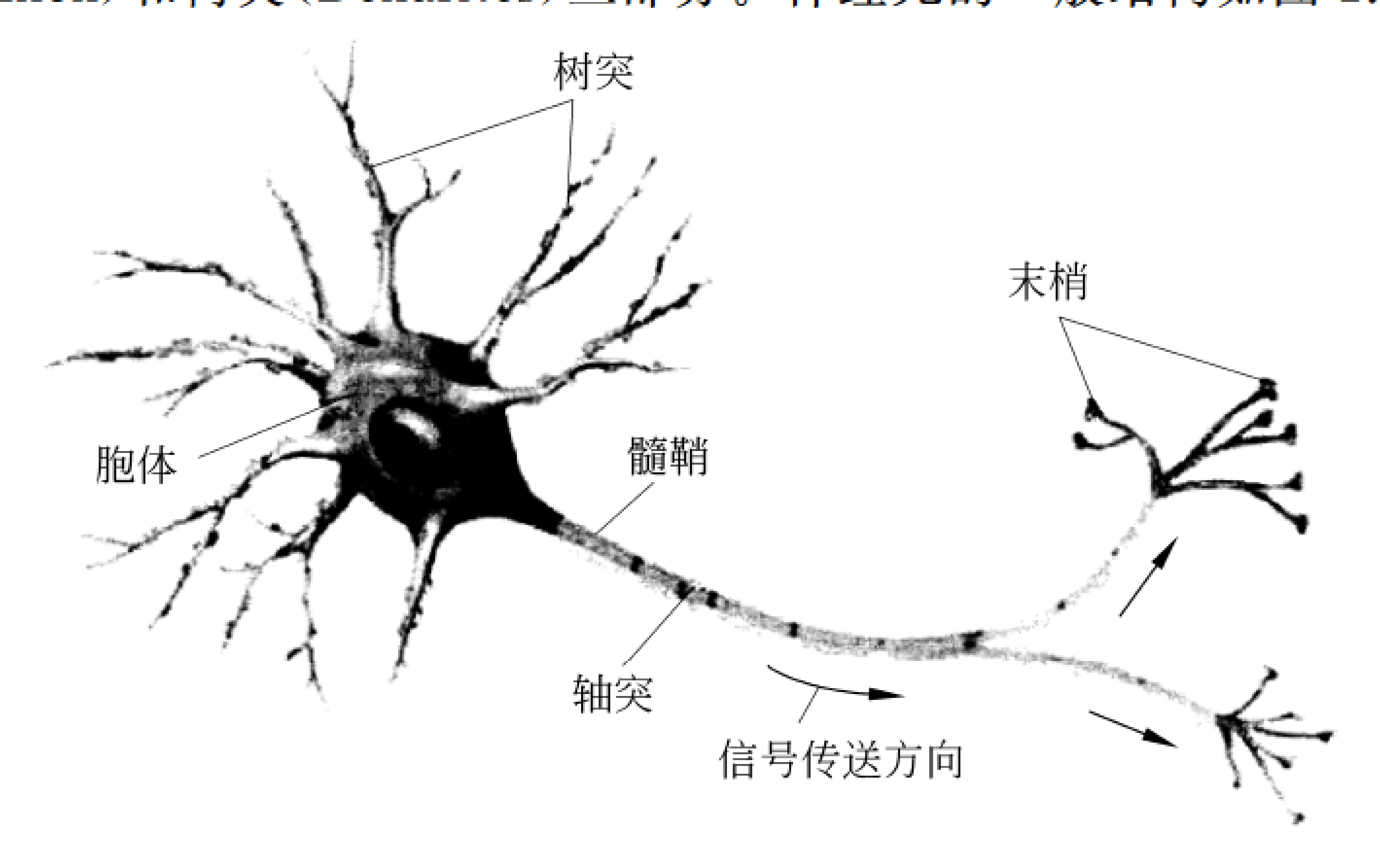



图 2.2　神经元的形态

胞体(Soma or Cell Body)是神经元的主体,位于脑和脊髓的灰质及神经节内,其形态各异,常见的形态为星形、锥体形、梨形和圆球形等。胞体大小不一,直径为 5～150μm。胞体是神经元的代谢和营养中心。胞体的结构与一般细胞相似,有核仁、细胞膜、细胞质和细胞核。胞内原浆在活细胞内呈颗粒状,经固定染色后显示内含神经原纤维、核外染色质(尼氏体、高尔基氏体、内质网和粒线体等)。神经原纤维是神经元特有的。

胞体的胞膜和突起表面的膜是连续完整的细胞膜。除突触部位的胞膜有特优的结构外,大部分胞膜为单位膜结构。神经细胞膜的特点是一个敏感而易兴奋的膜。在膜上有各种受体(Receptor)和离子通道(Ionic Chanel),二者各由不同的膜蛋白所构成。形成突触部分的细胞膜增厚。膜上受体可与相应的化学物质神经递质结合。当受体与乙酰胆碱递质或γ-氨基丁酸递质结合时,膜的离子通透性及膜内外电位差发生改变,胞膜产生相应的生理活动——兴奋或抑制。

细胞核多位于神经细胞体中央,大而圆,异染色质少,多位于核膜内侧,常染色质多,散在于核的中部,故着色浅,核仁 1～2 个,大而明显。细胞变性时,核多移向周边而偏位。

细胞质位于核的周围,又称核周体(Perikaryon)其中含有发达的高尔基复合体、滑面内质网,丰富的线粒体、尼氏体及神经原纤维,还含有溶酶体、脂褐素等结构。具有分泌功能的神经元,胞质内还含有分泌颗粒,如位于下丘脑的一些神经元。

神经元的突起是神经元胞体的延伸部分,由于形态结构和功能的不同,可分为树突和轴突。

(1) 树突(Dendrite):是从胞体发出的一至多个突起,呈放射状。胞体起始部分较粗,经反复分支而变细,形如树枝状。树突的结构与胞体相似,胞质内含有尼氏体,线粒体和平行排列的神经原纤维等,但无高尔基复合体。在特殊银染标本上,树突表面可见许多棘状突起,长 0.5～1.0μm,粗 0.5～2.0μm,称树突棘(Dendritic Spine),是形成突触的部位。一般电镜下,树突棘内含有数个扁平的囊泡称棘器(Spine Apparatus)。树突的分支和树突棘可扩大神经元接受刺激的表面积。树突具有接受刺激并将冲动传入细胞体的功能。

(2) 轴突(Axon):每个神经元只有一根,它在胞体上发出的轴突多呈锥形,称轴丘(Axon Hillock),其中没有尼氏体,主要有神经原纤维分布。轴突自胞体伸出后开始的一段,称为起始段,长 15～25μm,通常较树突细,粗细均匀,表面光滑,分支较少,无髓鞘包卷。离开胞体一定距离后,有髓鞘包卷,即为有髓神经纤维。轴突末端多呈纤细分支称轴突终末(Axon Terminal),与其他神经元或效应细胞接触。轴突表面的细胞膜,称轴膜(Axolemma),轴突内的胞质称轴质(Axoplasm)或轴浆。轴质内有许多与轴突长袖平行的神经原纤维和细长的线粒体,但无尼氏体和高尔基复合体,因此,轴突内不能合成蛋白质。轴突成分代谢更新以及突触小泡内神经递质,均在胞体内合成,通过轴突内微管、神经丝流向轴突末端。轴突的主要功能是将神经冲动由胞体传至其他神经元或效应细胞。轴突传导神经冲动的起始部位,是在轴突的起始段,沿轴膜进行传导。轴突的末梢,经连续分枝,以球形膨大的梢足与其他神经细胞或效应器细胞构成突触(Synapse)联系。

在长期的进化过程中,神经元在各自的机能和形态上都特化了。直接与感受器相联系把信息传向中枢的成为感觉神经元,或称传入神经元。直接与效应器相联系,把冲动从中枢传到效应器的成为运动神经元,或称传出神经元。除了上述传入传出神经元外,其余大量的神经元都是中间神经元,它们形成神经网络。

人体中枢神经系统的传出神经元的数目总计为数十万。传入神经元较传出神经元多1～3 倍。而中间神经元的数目最大，单就以中间神经元组成的大脑皮质来说，一般认为有140 亿～150 亿。

## 2.2.2　神经元的分类

神经元的分类有多种方法，常以神经元突起的数目和功能进行分类。

### 1. 按神经元突起的数目分类

根据神经元突起的数目，可将其分为以下三类。

（1）假单极神经元（Pseudounipolar Neuron）：从胞体发出一个突起，在离胞体不远处呈 T 型分为两支，因此，称假单极神经元。其中一支突起细长，结构与轴突相同，伸向周围，称周围突（Peripheral Process），其功能相当于树突，能感受刺激并将冲动传向胞体；另一分支伸向中枢，称中枢突（Central Process），将冲动传给另一个神经元，相当于轴突。如脊神经节内的感觉神经元等。

（2）双极神经元（Bipolar Neuron）：从胞体两端各发出一个突起，一个是树突，另一个是轴突。如耳蜗神经节内的感觉神经元等。

（3）多极神经元（Multipolar Neuron）：有一个轴突和多个树突，是人体中数量最多的一种神经元，如脊髓前角运动神经元和大脑皮质的锥体细胞等。多极神经元又可依轴突的长短和分支情况分为两型：高尔基Ⅰ型神经元，其胞体大，轴突长，在行径途中发出侧支，如脊髓前角运动神经元；高尔基Ⅱ型神经元，其胞体小，轴突短，在胞体附近发出侧支，如脊髓后角的小神经元以及大、小脑内的联合神经元。

### 2. 按神经元功能分类

根据神经元的功能，可将其分为以下三类。

（1）感觉神经元：也称传入神经元（Afferent Neuron），是传导感觉冲动的，胞体在脑、脊神经节内，多为假单极神经元。其突起构成周围神经的传入神经。神经纤维终末在皮肤和肌肉等部位形成感受器。

（2）运动神经元：也称传出神经元，是传导运动冲动的神经元，多为多极神经元。胞体位于中枢神经系统的灰质和植物神经节内，其突起构成传出神经纤维。神经纤维终末，分布在肌组织和腺体，形成效应器。

（3）中间神经元：也称联合神经元是在神经元之间起联络作用的神经元，是多极神经元，人类神经系统中，最多的神经元，构成中枢神经系统内的复杂网络。胞体位于中枢神经系统的灰质内，其突起一般也位于灰质。

## 2.2.3　神经胶质细胞

神经胶质细胞或简称胶质细胞（Glial Cell），广泛分布于中枢和周围神经系统，其数量比神经元的数量大得多，胶质细胞与神经元数目之比为 10∶1～50∶1。胶质细胞与神经元一样具有突起，但其胞突不分树突和轴突，亦没有传导神经冲动的功能。胶质细胞可分几

种,各有不同的形态特点。

### 1. 星形胶质细胞

星形胶质细胞(Astrocyte)是胶质细胞中体积最大的一种,与少突胶质细胞合称为大胶质细胞(Macroglia)。细胞呈星形,核圆形或卵圆形,较大,染色较浅(见图 2.3)。

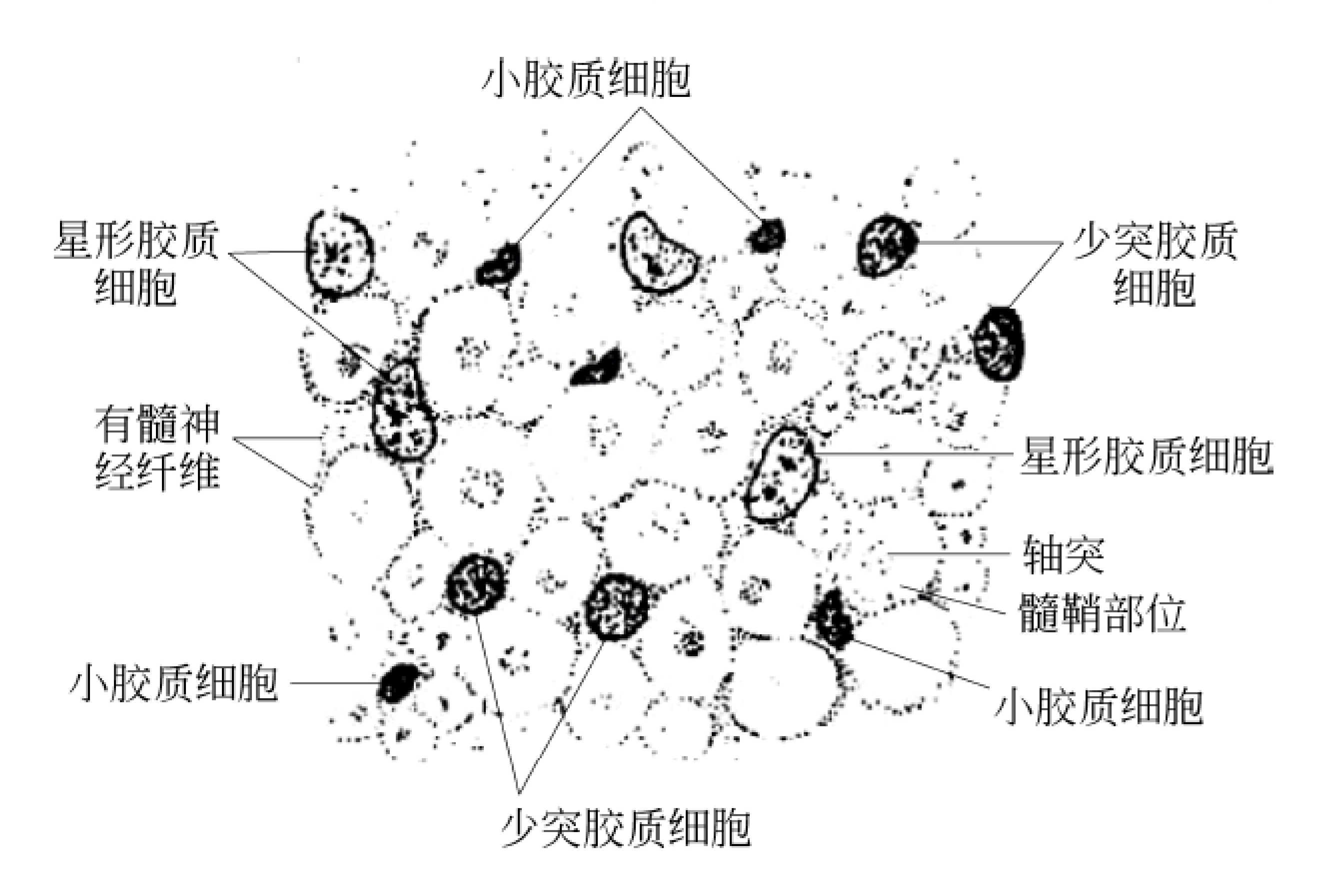



图 2.3 神经胶质细胞的细胞核及神经纤维横切

(脊髓白质,Nissl 法染色)

星形胶质细胞可分两种:

(1) 纤维性星形胶质细胞(Fibrous Astrocyte),多分布在白质,细胞的突起细长,分支较少,胞质内含大量胶质丝(Glial Fiament)。组成胶质丝的蛋白质称胶质原纤维酸性蛋白(Glial Fibrillary Acidic Protein,GFAP),用免疫细胞化学染色技术,能特异性地显示这类细胞。

(2) 原浆性星形胶质细胞(Protoplasmic Astrocyte),多分布在灰质,细胞的突起较短粗,分支较多,胞质内胶质丝较少。星形胶质细胞的突起伸展充填在神经元胞体及其突起之间,起支持和分神经元的作用。有些突起末端形成脚板,附在毛细血管壁上,或附着在脑和脊髓表面形成胶质界膜。

星形胶质细胞之间的细胞间隙狭窄而迂回曲折,宽 15～20nm,内含组织液,神经元借此进行物质交换(图 2.4),星形胶质细胞能吸收细胞间隙的 $K^+$,以维持神经元周围环境 $K^+$ 含量的稳定性,它还能摄取和代谢某些神经递质(如 γ-氨基丁酸等),调节细胞间隙中神经递质的浓度,有利神经元的活动。在神经系统发育时期,某些星形胶质细胞具有引导神经元迁移的作用,使神经元到达预定区域并与其他细胞建立突触连接。中枢神经系统损伤时,星形胶质细胞增生、肥大、充填缺损的空隙,形成胶质瘢痕(Glial Scar)。

### 2. 少突胶质细胞

在银染色标本中,少突胶质细胞(Oligodendrocyte)的突起较少,但用特异性的免疫细胞化学染色,可见少突胶质细胞的突起并不少,而且分支也多。少突胶质细胞的胞体较星形胶质细胞的小,核圆,染色较深(图 2.3)。胞质内胶质丝很少,但有较多微管和其他细胞器。少突胶质细胞分布在神经元胞体附近和神经纤维周围,它的突起末端扩展成扁平薄膜,包卷神经元的轴突形成髓鞘,所以它是中枢神经系统的髓鞘形成细胞。新近研究认为,少突胶质

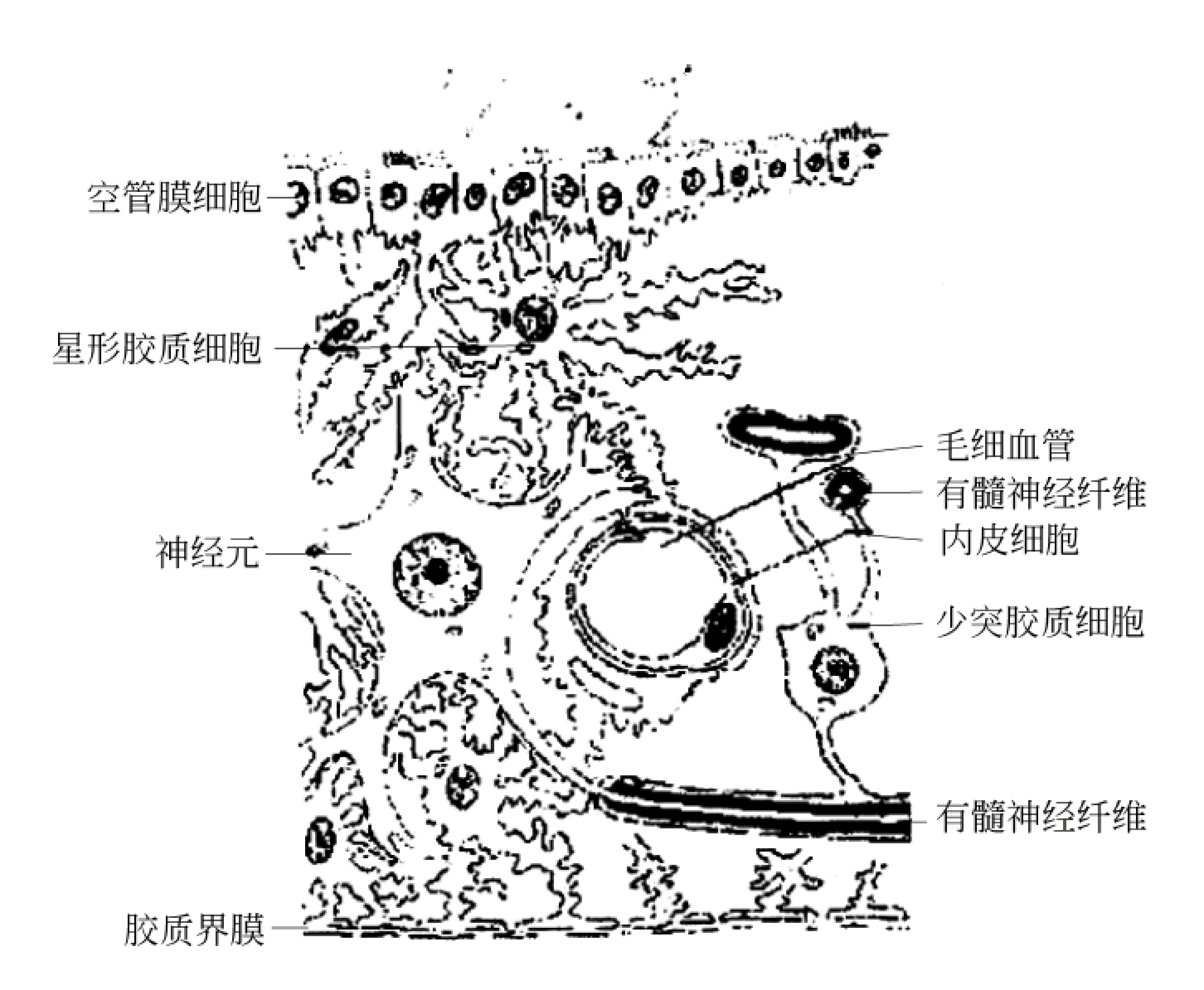



图 2.4 神经胶质细胞与神经元和毛细胞血管

细胞还有抑制再生神经元突起生长的作用。

### 3. 小胶质细胞

小胶质细胞(Microglia)是胶质细胞中最小的一种。胞体细长或椭圆,核小,扁平或三角形,染色深(见图 2.3)。细胞的突起细长有分支,表面有许多小棘突。小胶质细胞的数量少,约占全部胶质细胞的 5%左右。中枢神经系统损伤时,小胶质细胞可转变为巨噬细胞,吞噬细胞碎屑及退化变性的髓鞘。血循环中的单核细胞亦侵入损伤区,转变为巨噬细胞,参与吞噬活动。由于小胶质细胞有吞噬功能,有人认为它来源于血液中的单核细胞,属单核吞噬细胞系统。

### 4. 室管膜细胞

室管膜细胞(Ependymal Cell)为立方或柱形,分布在脑室及脊髓中央管的腔面,形成单层上皮,称室管膜(Ependyma)。室管膜细胞表面有许多微绒毛,有些细胞表面有纤毛(见图 2.4)。某些地方的室管膜细胞,其基底面有细长的突起伸向深部,称伸长细胞(Tanycyte)。

## 2.3 突触传递

神经元与神经元之间,或神经元与非神经细胞(肌细胞、腺细胞等)之间的一种特化的细胞连接,称为突触(Synapse)[63]。它是神经元之间的联系和进行生理活动的关键性结构。通过它的传递作用实现细胞与细胞之间的通信。在神经元之间的连接中,最常见是一个神经元的轴突终末与另一个神经元的树突、树突棘或胞体连接,分别构成轴-树、轴-棘、轴-体突触。此外还有轴-轴和树-树突触等。突触可分为化学突触和电突触两大类。前者是以化学物质(神经递质)作为通信的媒介,后者是亦即缝隙连接,是以电流(电信号)传递信息。哺乳

动物神经系统以化学突触占大多数,通常所说的突触是指化学突触。

突触的结构可分突触前成分、突触间隙和突触后成分三部分。突触前、后成分彼此相对的细胞膜分别称为突触前膜和突触后膜,两者之间宽 15～30nm 的狭窄间隙为突触间隙,内含糖蛋白和一些细丝。突触前成分通常是神经元的轴突终末,呈球状膨大,附着在另一神经元的胞体或树突上,称突触扣结。

## 2.3.1 化学性突触

电镜下,突触扣结内含许多突触小泡,还有少量线粒体、滑面内质网、微管和微丝等(图 2.5)。突触小泡的大小和形状不一,多为圆形,直径为 40～60nm,亦有的呈扁平形。突触小泡有的清亮,有的含有致密核芯(颗粒型小泡),大的颗粒型小泡直径可达 200nm。突触小泡内含神经递质或神经调质。突触前膜和后膜均比一般细胞膜略厚,这是由于其胞质面附有一些致密物质所致(图 2.5)。在突触前膜还有电子密度高的锥形致密突起(Dense Projection)突入胞质内,突起间容纳突触小泡。突触小泡表面附有突触小泡相关蛋白,称突触素Ⅰ(SynapsinⅠ),它使突触小泡集合并附在细胞骨架上。突触前膜上富含电位门控通道,突触后膜上则富含受体及化学门控通道。当神经冲动沿轴膜传至轴突终末时,即触发突触前膜上的电位门控钙通道开放,细胞外的 $Ca^{2+}$ 进入突触前成分,在 ATP 的参与下使突触素Ⅰ发生磷酸化,促使突触小泡移附在突触前膜上,通过出胞作用释放小泡内的神经递质到突触间隙内。其中部分神经递质与突触后膜上相应受体结合,引起与受体耦联的化学门控通道开放,使相应离子进出,从而改变突触后膜两侧离子的分布状况,出现兴奋或抑制性变化,进而影响突触后神经元(或非神经细胞)的活动。使突触后膜发生兴奋的突触称为兴奋性突触,使突触后膜发生抑制的称为抑制性突触。

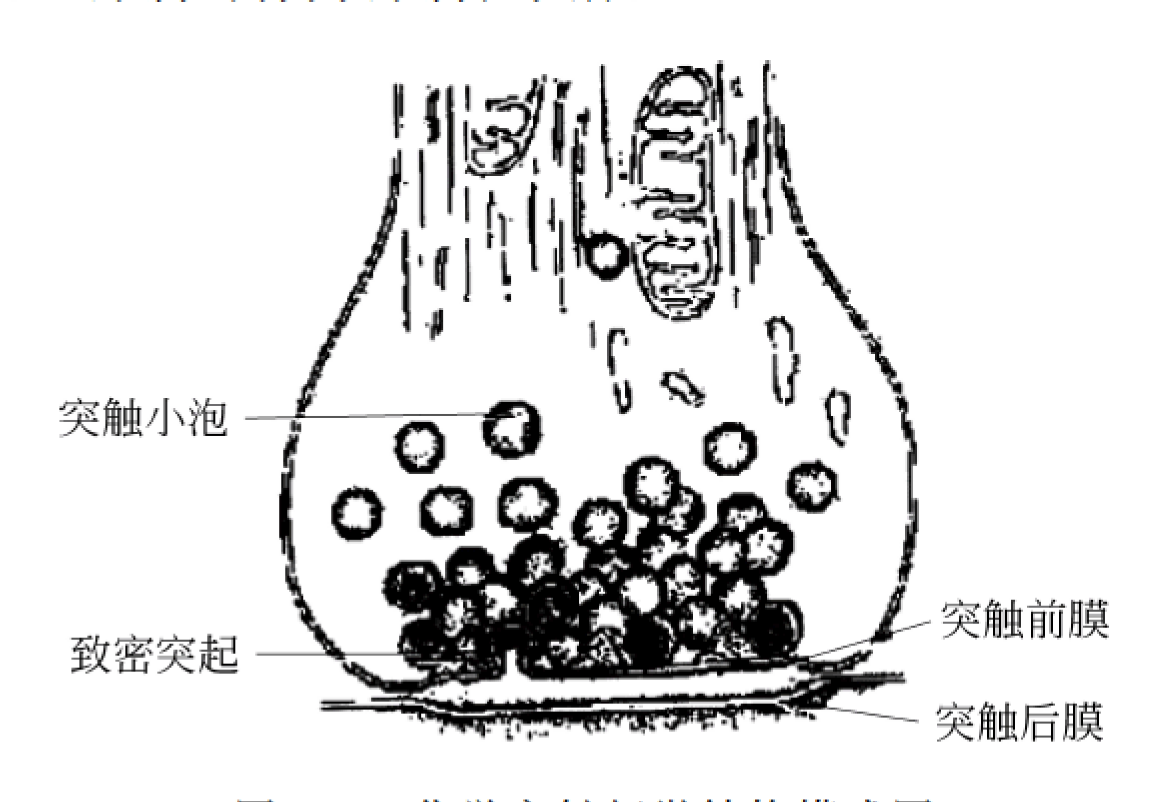



图 2.5 化学突触超微结构模式图

### 1. 突触前部

神经元轴突终末呈球状膨大,轴膜增厚形成突触前膜,突触厚 6～7nm。在突触前膜部位的胞浆内,含有许多突触小泡以及一些微丝和微管、线粒体和滑面内质网等。突触小泡是突触前部的特征性结构,小泡内含有化学物质,称为神经递质。各种突触内的突触小泡形状和大小颇不一致,是因其所含神经递质不同。常见突触小泡类型有如下三种。

(1) 球形小泡,直径 20～60nm,小泡清亮,其中含有兴奋性神经递质,如乙酰胆碱。

（2）颗粒小泡，小泡内含有电子密度高的致密颗粒，按其颗粒大小又可分为两种：小颗粒小泡直径30～60nm，通常含胺类神经递质如肾上腺素、去甲肾上腺素等；大颗粒小泡直径可达80～200nm，所含的神经递质为5-羟色胺或脑啡肽等肽类。

（3）扁平小泡，小泡长径约50nm，呈扁平圆形，其中含有抑制性神经递质，如γ-氨基丁酸等。

各种神经递质在胞体内合成，形成小泡，通过轴突的快速顺向运输到轴突末端。新近研究发现在中枢和周围神经系统中，有两种或两种以上神经递质共存于一个神经元中，在突触小体内可有两种或两种以上不同形态的突触小泡。如交感神经节内的神经细胞，有乙酸胆碱和血管活性肠肽。前者支配汗腺分泌；后者作用于腺体周围的血管平滑肌使其松弛，增加局部血流量。神经递质共存的生理功能，是协调完成神经生理活动作用，使神经调节更加精确和协调。目前，许多事实表明，递质共存不是个别现象，而是一个普遍性规律，有许多新的共存递质和新的共存部位已被证实。其中多为非肽类递质（胆碱类、单胺类和氨基酸类）和肽类递质共存。

关于突触小泡的包装、储存和释放递质的问题，现已知突触体素，突触素和小泡相关膜蛋白等三种蛋白与之有关。突触体素是突触小泡上 $Ca^{2+}$ 的结合蛋白，当兴奋剂到达突触时，$Ca^{2+}$ 内流突然增加而与这种蛋白质结合，可能对突触小泡的胞吐起重要作用。突触素是神经细胞的磷酸蛋白，有调节神经递质释放的作用，小泡相关膜蛋白（VAMP）是突触小泡膜的结构蛋白，可能对突触小泡代谢有重要作用。

### 2. 突触后部

多为突触后神经元的胞体膜或树突膜，与突触前膜相对应部分增厚，形成突触后膜。厚为20～50nm，比突触前膜厚，在后膜具有受体和化学门控的离子通道。根据突触前膜和后膜的胞质面致密物质厚度不同，可将突触分为Ⅰ和Ⅱ两型：Ⅰ型突触后膜胞质面致密物质比前膜厚，因而膜的厚度不对称，故又称为不对称突触；突触小泡呈球形，突触间隙较宽（20～50nm）；一般认为Ⅰ型突触是兴奋性突触，主要分布在树突干上的轴-树突触。Ⅱ型突触前、后膜的致密物质较少，厚度近似，故称为对称性突触，突触小泡呈扁平形，突触间隙也较窄（10～20nm）。认为Ⅱ型突触是一种抑制性突触，多分布在胞体上的轴-体突触。

### 3. 突触间隙

突触间隙是位于突触前、后膜之间的细胞外间隙，宽20～30nm，其中含糖胺多糖（如唾液酸）和糖蛋白等，这些化学成分能和神经递质结合，促进递质由前膜移向后膜，使其不向外扩散或消除多余的递质。

突触的传递过程，是神经冲动沿轴膜传至突触前膜时，触发前膜上的电位门控钙通道开放，细胞外的 $Ca^{2+}$ 进入突触前部，在ATP和微丝、微管的参与下，使突触小泡移向突触前膜，以胞吐方式将小泡内的神经递质释放到突触间隙。其中部分神经递质与突触后膜上的相应受体结合，引起与受体偶联的化学门控通道开放，使相应的离子经通道进入突触后部，使后膜内外两侧的离子分布状况发生改变，呈现兴奋性（膜的去极化）或抑制性（膜的极化增强）变化，从而影响突触后神经元（或效应细胞）的活动。使突触后膜发生兴奋的突触，称为兴奋性突触，而使后膜发生抑制的称为抑制性突触。

突触的兴奋或抑制决定于神经递质及其受体的种类,神经递质的合成、运输、储存、释放、产生效应以及被相应的酶作用而失活,是一系列神经元的细胞器生理活动。一个神经元通常有许多突触,其中有些是兴奋性的,有些是抑制性的。如果兴奋性突触活动总和超过抑制性突触活动总和,并达到能使该神经元的轴突起始段发生动作电位,出现神经冲动时,则该神经元呈现兴奋,反之,则表现为抑制。

化学突触的特征,是一侧神经元通过出胞作用释放小泡内的神经递质到突触间隙,相对应一侧的神经元(或效应细胞)的突触后膜上有相应的受体。具有这种受体的细胞称为神经递质的效应细胞或靶细胞,这就决定了化学突触传导为单向性。突触的前后膜是两个神经膜特化部分,维持两个神经元的结构和功能,实现机体的统一和平衡。故突触对内、外环境变化很敏感,如缺氧、酸中毒、疲劳和麻醉等,可使兴奋性降低。茶碱、碱中毒等则可使兴奋性增高。

## 2.3.2 电突触

电突触是神经元间传递信息的最简单形式,在两个神经元间的接触部位,存在缝隙连接,接触点的直径在 0.1～10μm 以上。也有突触前、后膜及突触间隙。突触的结构特点,突触间隙仅 1～1.5nm,前、后膜内均有膜蛋白颗粒,显示呈六角形的结构单位,跨跃膜的全层,顶端露于膜外表,其中心形成一微小通道,此小管通道与膜表面相垂直,直径约为 2.5nm,小于 1nm 的物质可通过,如氨基酸。缝隙连接两侧膜是对称的。相邻两突触膜,膜蛋白颗粒顶端相对应,直接接触,两侧中央小管,由此相通。轴突终末无突触小泡,传导不需要神经递质,是以电流传递信息,传递神经冲动一般均为双向性。神经细胞间电阻小,通透性好,局部电流极易通过。电突触功能有双向快速传递的特点,传递空间减少,传送更有效。

现在已证明,哺乳动物大脑皮质的星形细胞,小脑皮质的篮状细胞、星形细胞,视网膜内水平细胞、双极细胞,以及某些神经核,如动眼神经运动核前、庭神经核、三叉神经脊束核,均有电突触分布。电突触的形式多样,可见有树-树突触、体-体突触、轴-体突触、轴-树突触等。

电突触对内、外环境变化很敏感。在疲劳、缺氧、麻醉或酸中毒情况下,可使兴奋性降低。而在碱中毒时,可使兴奋性增高。

## 2.3.3 突触传递的机制

突触传递的基本过程是动作电位传到轴突末梢,引起小体区域的去极化,增加 $Ca^{2+}$ 的通透性,细胞外液的 $Ca^{2+}$ 流入,促使突触小泡前移与突触前膜融合,在融合处出现破口,使池内所含的介质释放到突触间隙,弥散与突触后膜特异性受体结合。然后化学门控性通道开放,突触后膜对某些离子通透性增加,突触后膜电位变化(突触后电位)(去极化或超极化),产生总和效应,引起突触后神经元兴奋或抑制。图 2.6 给出突触传递的基本过程简单示意图。

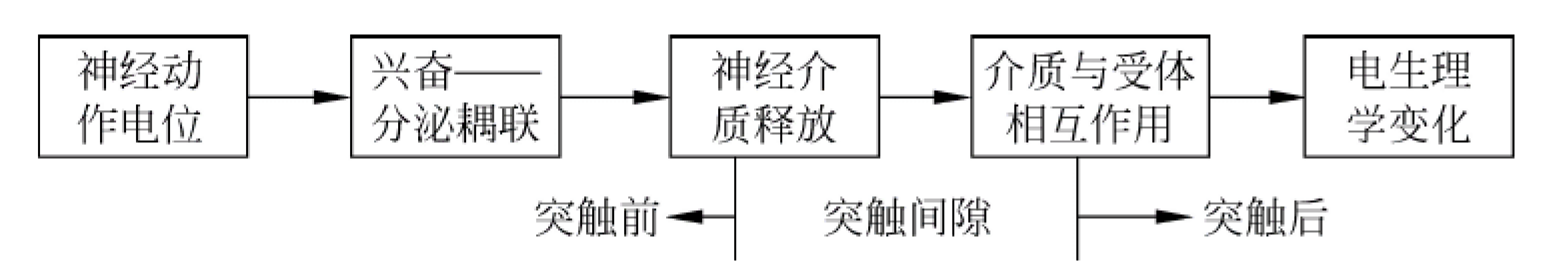



图 2.6 突触传递的基本过程简单示意图

$Ca^{2+}$ 在突触传递中的作用如下。

(1) 降低轴浆的黏度，有利于突触小泡的位移(降低囊泡上肌动蛋白结合蛋白与肌动蛋白的结合)。

(2) 消除突触前膜内侧的负电位，促进突触小泡和前膜接触、融合和胞裂，促进神经递质的释放。

在高等动物神经系统突触前的电活动，从不直接引起突触后成分的活动，不存在电学耦联。突触传递一律通过特殊的化学物质中介，这种物质就叫做神经介质或递质。突触传递只能由突触前到突触后，在这个系统中不存在反方向活动的机制。因此突触传递是单方向的。这里兴奋——分泌的耦联(介质释放)和介质在间隙的扩散，直到突触后膜的去极化需 0.5～1ms，这就是突触迟延。突触传递具有如下特征。

(1) 单向传递(因为只有前膜能释放递质)。

(2) 突触迟延。

(3) 总和，包括时间性总和与空间性总和。

(4) 对内环境变化敏感和易疲劳。

(5) 兴奋节律性改变(同一反射活动中传入神经与传出神经发放的频率不一致)。

(6) 后放(刺激停止后，传出神经在一定时间内仍发放冲动)。

## 2.4　神经递质

英国剑桥大学医学院的学生埃利奥特(T R Elliott)于 1904 年在生理学会宣读的论文中曾谨慎地指出：冲动传到交感神经末梢，可能是从那里释放肾上腺素、再作用到效应器细胞。这是明确地指出化学传递的最早记录。于 1921 年首先给化学传递以实验证明的是奥地利生理学家勒韦(O Loewi)。他以电刺激离体灌流蛙心标本的迷走神经时观察到，在心跳受到抑制同时，在灌流液中出现了可抑制另一个灌流蛙心跳动的物质，并把它称为迷走素。于 1926 年他又进一步阐明了迷走素即是 Ach。这是第一个被发现的递质。英国生理学家和药理学家戴尔(H H Dale)和同事曾提出实验证据表明，Ach 是骨骼肌神经肌肉接头的递质。他还把胆碱能(Cholinergic)一词引入生理学，用以表示以 Ach 为递质的神经元。勒韦和戴尔因为这项研究于 1936 年获得了诺贝尔奖。关于交感神经末梢可释放肾上腺素的设想虽最早被提出，在 1921 年坎农(W B Cannon)还将刺激交感神经从肝脏释放的物质命名为交感素(Sympathin)，并认为它虽与肾上腺素十分相似，但也有差异，直到 1949 年该物质方被奥伊勒(von Euler)鉴定为去甲肾上腺素(Noradrenaline，NA)，为此他获得了 1970 年度诺贝尔奖。此后，又相继发现了为数不多的(约 10 种)小分子递质和多种(50 种以上)参与突触传递的神经活性多肽，以及近年又发现了一氧化氮(NO)为气体信使[63]。

关于判断内源性神经活性物质是否为递质，一般有五条鉴定标准。

(1) 存在。应特异性地存在于以该物质为递质的神经元中，而且，在这种神经元的末梢有合成该递质的酶系统。

(2) 部位。递质在神经末梢内合成以后，通常是集中储存在囊泡(Vesicle)内，这样可以防止被胞浆内的其他酶所破坏。

(3) 释放。从突触前末梢可释放足以在突触后细胞或效应器引起一定反应的物质。

(4) 作用。递质通过突触间隙,作用于突触后膜的叫做受体的特殊部位,引起突触后膜离子通透性改变以及电位变化。

(5) 灭活机制。神经递质在发挥上述效应后,其作用应该迅速终止,以保证突触传递的高度灵活。作用的终止有几种方式:一是被酶所水解,失去活性;其次是被突触前膜"重摄取",或是一部分为后膜所摄取;也有的部分进入血循环,在血中一部分被酶所降解破坏。

目前已知的神经递质种类很多,但主要的有乙酰胆碱、儿茶酚胺类(去甲肾上腺素和多巴胺)、5-羟色胺、GABA、某些氨基酸和寡肽等。

## 2.4.1 乙酰胆碱

乙酰胆碱(Acetylcholine,简写为 Ach)是许多外周神经如运动神经、植物性神经系统的节前纤维和付交感神经节后纤维的兴奋性神经递质。

Ach 由胆碱和乙酰 CoA 所合成。胆碱乙酰化酶(Choline Acetylase)催化下列反应:

$$\underset{\text{胆碱}}{(CH_3)_3N^+—CH_2—CH_2—OH}+\underset{\text{乙酰辅酶 A}}{CH_3—CO\sim CoA}\xrightarrow{\text{胆碱乙酰化酶}}$$

$$\underset{\text{乙酰胆碱}}{(CH_3)_3N^+—CH_2—CH_2—O—CO—CH_3}+\underset{\text{辅酶 A}}{CoA}$$

由于胆碱乙酰化酶位于胞浆内,因此设想 Ach 是先在胞浆内合成,然后进入囊泡储存。平时囊泡中和胞浆中的 Ach 大约各占一半,且两者可能处于平衡状态。囊泡内储存的 Ach 是一种结合型的(与蛋白质结合),而释放至胞浆时,则变为游离型。

当神经冲动沿轴突到达末梢时,囊泡趋近突触膜,并与之融合、破裂,此时囊泡内结合型 Ach 转变为游离型 Ach,释放入突触间隙。同时,还可能有一部分胞浆内新合成的 Ach 也随之释放。

Ach 作用于突触后膜(突触后神经元或效应细胞的膜)表面的受体,引起生理效应。已经确定 Ach 受体是一种分子量为 42 000 的蛋白质,通常以脂蛋白的形式存在于膜上。

Ach 在传递信息之后和受体分开,游离于突触间隙,其中极少部分在突触前膜的载体系统作用下重新被摄入突触前神经元。大部分 Ach 是在胆碱酯酶的作用下水解成胆碱和乙酸而失去活性,也有一部分经弥散而离开突触间隙。关于乙酰胆碱在神经末梢中的代谢,可参见图 2.7 和图 2.8。

## 2.4.2 儿茶酚胺类

儿茶酚胺类(Catecholamines)是指含有邻苯二酚基本结构的胺类。体内具有生物活性的儿茶酚胺包括多巴胺(Dopamine,DA)、去甲肾上腺素(Norepinephrine,Noradrenaline,NE)和肾上腺素(Epinephrine,Adrenalin,E)。它们的结构如下:

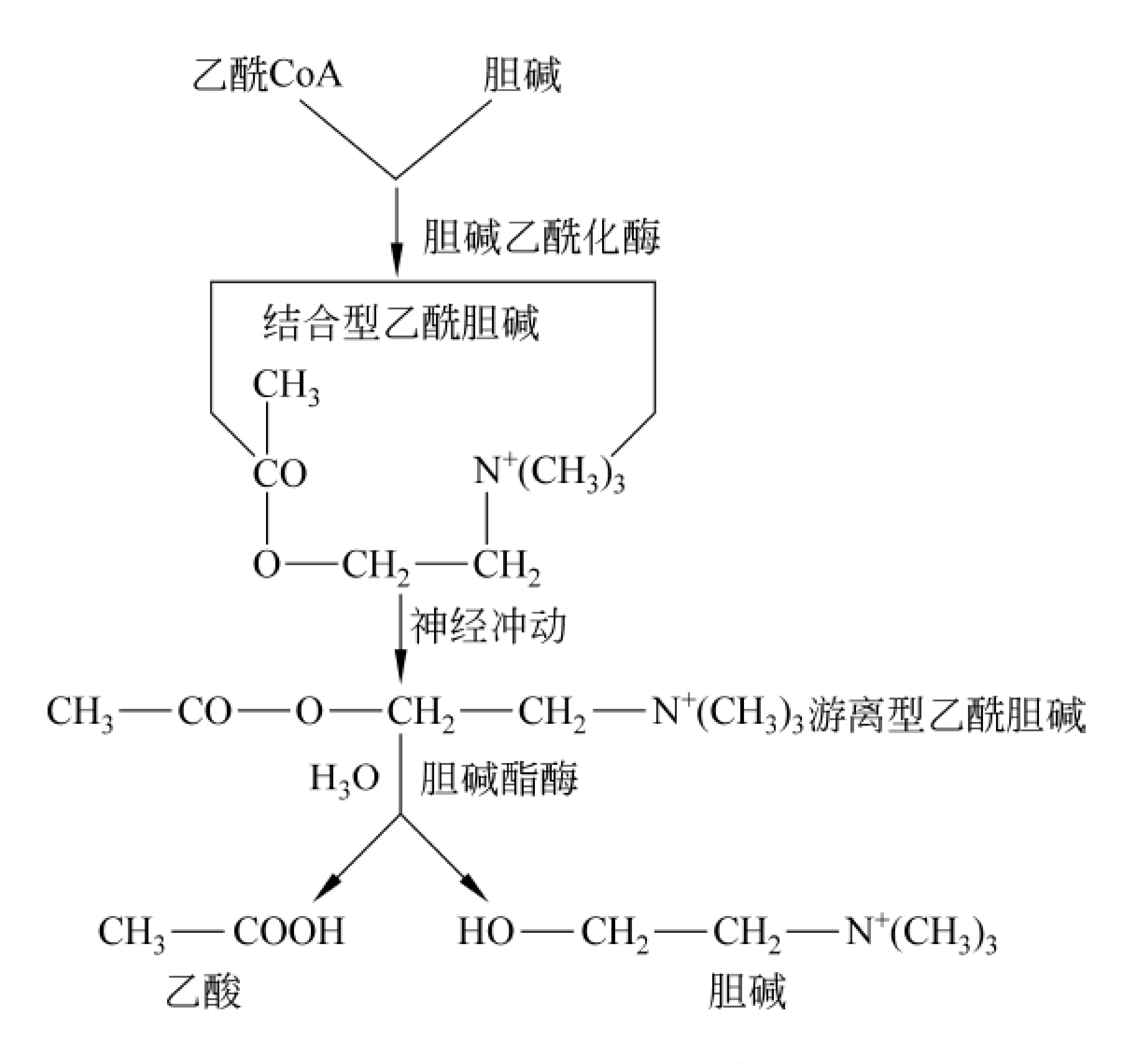



图 2.7　乙酰胆碱的代谢

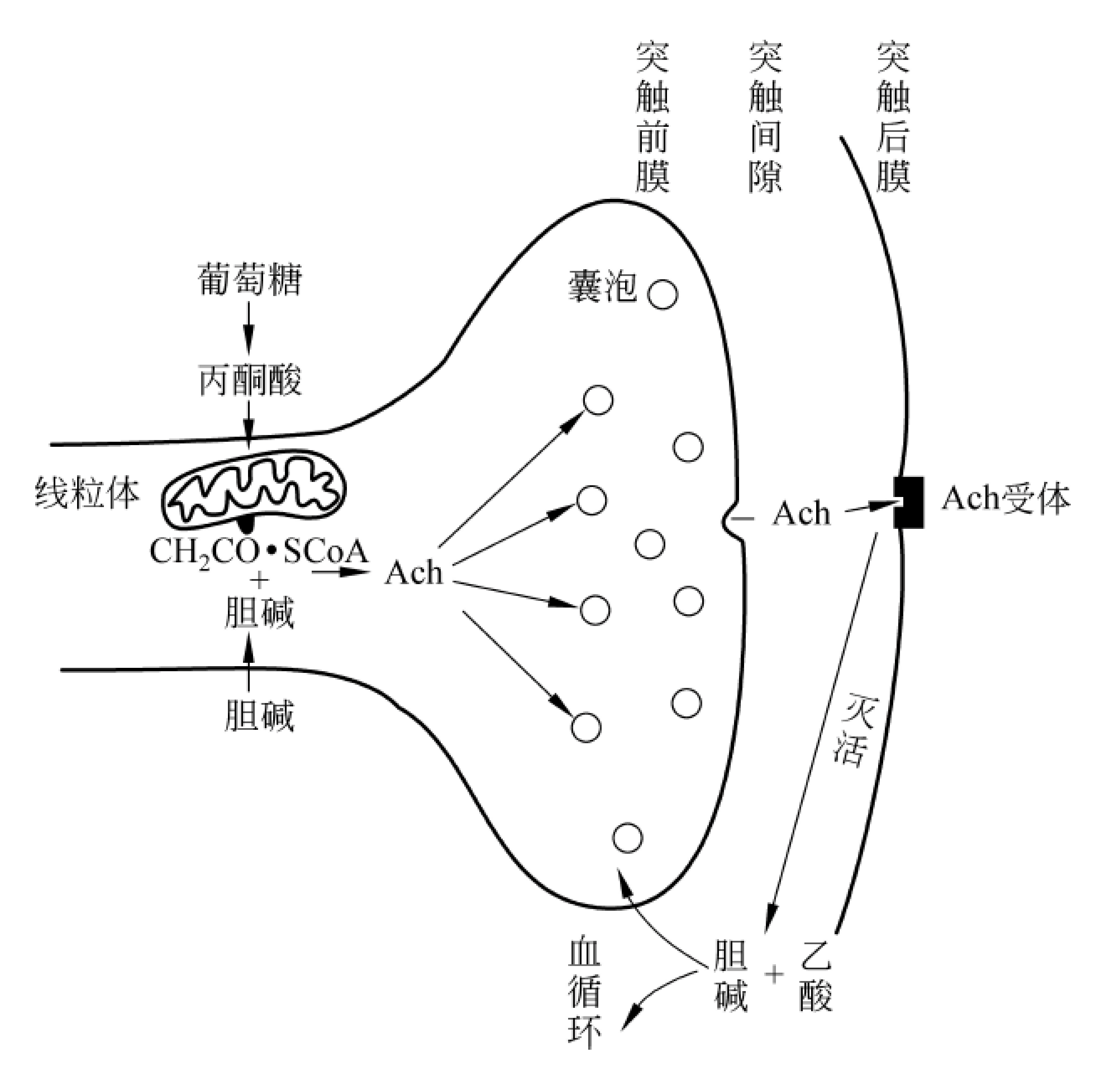



图 2.8　突触部位 Ach 的动态

去甲肾上腺素和肾上腺素既是肾上腺髓质所分泌的激素，又是交感和中枢神经系统中去甲肾上腺素能纤维的神经递质。NE 在中枢内分布广泛，含量较多，E 则少，因此我们着重介绍 NE 的代谢。DA 则主要集中在锥体外系，也是一种神经递质。

### 1. 儿茶酚胺的生物合成

神经组织中儿茶酚胺的合成原料来自血液中的酪氨酸，其合成过程如图 2.9 所示。

在上述过程中，第一步有酪氨酸羟化酶参加，它位于去甲肾上腺素能神经纤维的胞浆内，含量少，活性低，成为 NE 生成的限速酶，四氢生物蝶呤是它的辅酶，$O_2$ 和 $Fe^{++}$ 也是合

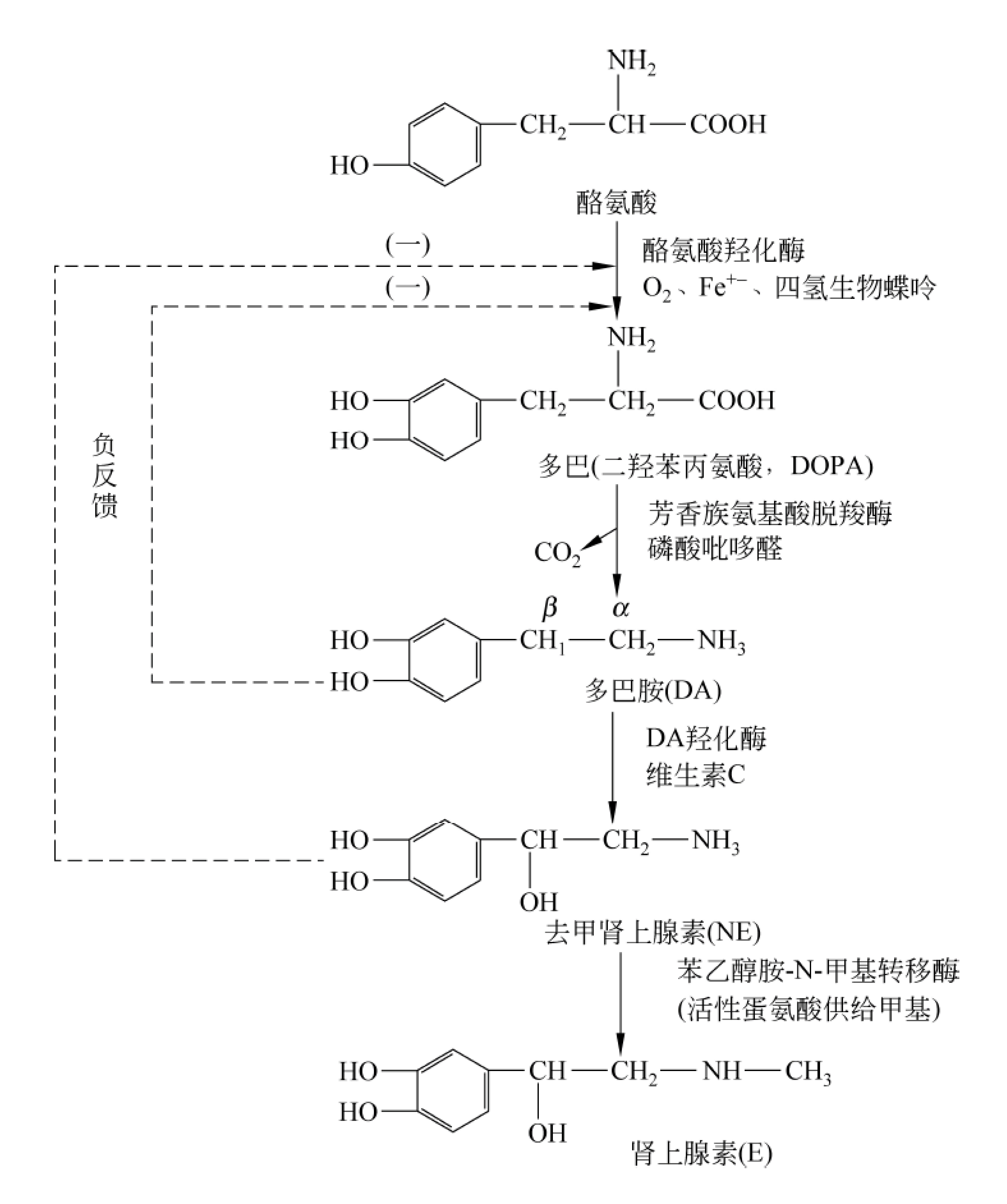



图 2.9 儿茶酚胺类的生物合成

成时必不可少的因素；第二步反应是由胞浆中芳香族氨基酸脱羧酶所催化，这酶的特异性不高，和一般氨基酸脱羧酶一样，要求磷酸吡哆醛作为辅酶；第三步是多巴胺羟化酶催化的氧化反应，氧化发生在 β 碳原子上，此酶不存在于胞浆而是附于囊泡内壁，属于含 $Cu^{++}$ 的蛋白质，并需要维生素 C 为辅助因子。

从以上各酶的亚细胞水平分布可以了解，合成 NE 的最后一步只能在囊泡内进行，NE 合成量不仅受酪氨酸羟化酶限速调节，而且当神经末梢胞浆中游离的 NE 浓度过高时，还可以因反馈性抑制酪氨酸羟化酶的作用而减少 NE 的合成。

苯乙醇胺-N-甲基转移酶主要见于肾上腺髓质细胞，可使 NE 甲基化生成肾上腺素。此酶在脑内虽有少量存在，但一般认为正常时哺乳类脑内的肾上腺素含量极少。有人认为，如果苯乙醇胺-N-甲基转移酶活性过高，可以使多巴胺直接变成 N-甲基多巴胺等物，造成这些递质的代谢紊乱，这可能是精神分裂症的原因之一。

### 2. 去甲肾上腺素(NE)

(1) 存储和释放：NE 在囊泡内合成后就地储存。它和 ATP、嗜铬颗粒蛋白等疏松地结合在一起，这样，就使 NE 不易渗入胞浆而遭单胺氧化酶的破坏。当神经冲动到达末梢时，突触前膜附近的囊泡便与前膜融合，破裂并生成小孔。这时囊泡内的 NE 连同嗜铬颗粒

蛋白等一起被释放入突触间隙。

(2) 突触间隙中 NE 的去向：释放入突触间隙的 NE 能与突触后膜上的 NE 受体结合，产生生理效应。然后，约有 3/4 的 NE 重新被突触前膜所摄取。摄入胞浆后的 NE 又可进入囊泡贮存。重摄取是一种耗能过程，与突触前膜上的 $Na^{+}$，$K^{+}$-ATP 酶系统及囊泡膜上的 $Mg^{++}$-ATP 酶系统有关。另外一部分 NE 则为后膜所摄取，并在那里被分解灭活。还有一部分是在突触间隙中被破坏或逸入血液。除了被突触前膜和囊泡重摄取的 NE 可再供利用外，其余的 NE 大都遭到酶促降解而灭活。NE 的酶促降解如图 2.10 所示，单胺氧化酶(MAO)和儿茶酚-O-甲基转移酶(Catechol-O-Transmethylase，COMT)是催化儿茶酚胺分解的两种主要的酶，它们不仅存在于神经组织内，而且广泛地分布于非神经组织，神经元的线粒体膜上也很多。NE 受 MAO 作用，首先氧化脱氨基生成醛，后者再变成醇或酸。3-甲氧基-4-羟基苯乙二醇(MHPG)是中枢内 NE 的主要降解产物。在外周则以氧化成香草基扁桃酸(VMA)为主。血液循环中的 NE(主要是激素)，则在肝、肾等组织经 COMT 作用变成甲氧基代谢产物而排泄。现在，临床上常测定尿中 VMA 含量，作为了解交感神经功能的指标，患嗜铬细胞瘤和神经母细胞瘤时，由于肿瘤组织也产生 NE 或 E，其代谢产物 VMA 相应增多，故在诊断上颇有意义。

关于 NE 在突触部位的动态可用图 2.11 表示。

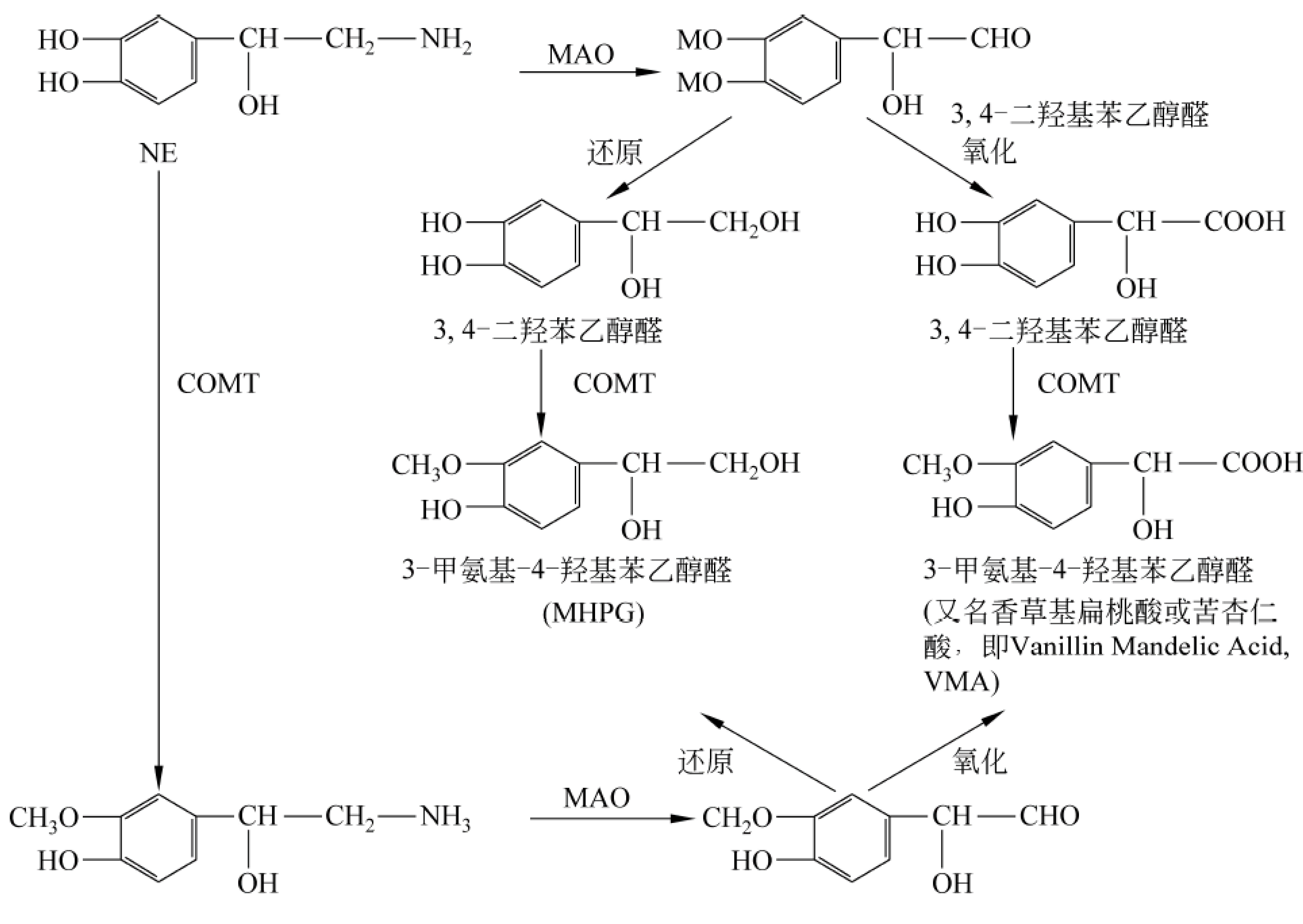



图 2.10　去甲肾上腺素的分解代谢

中枢神经系统内的去甲肾上腺素能神经元胞体集中在延脑和桥脑，其通路都已研究明确。但是 NE 在中枢究竟是抑制性还是兴奋性的递质，目前尚不能肯定，可能随部位不同而异。NE 所引起的生理效应也很难用简单的术语“兴奋”“抑制”来表达。根据一些动物实验观察到，NE 可引起动物嗜睡，体温降低，出现摄食行为。有人认为，脑内 NE 减少，可表现出精神抑郁；反之，过量可表现出狂躁。总之，脑内 NE 的功能可能和体温、摄食行为、镇

痛、心血管系统和精神状态的调节有密切关系。

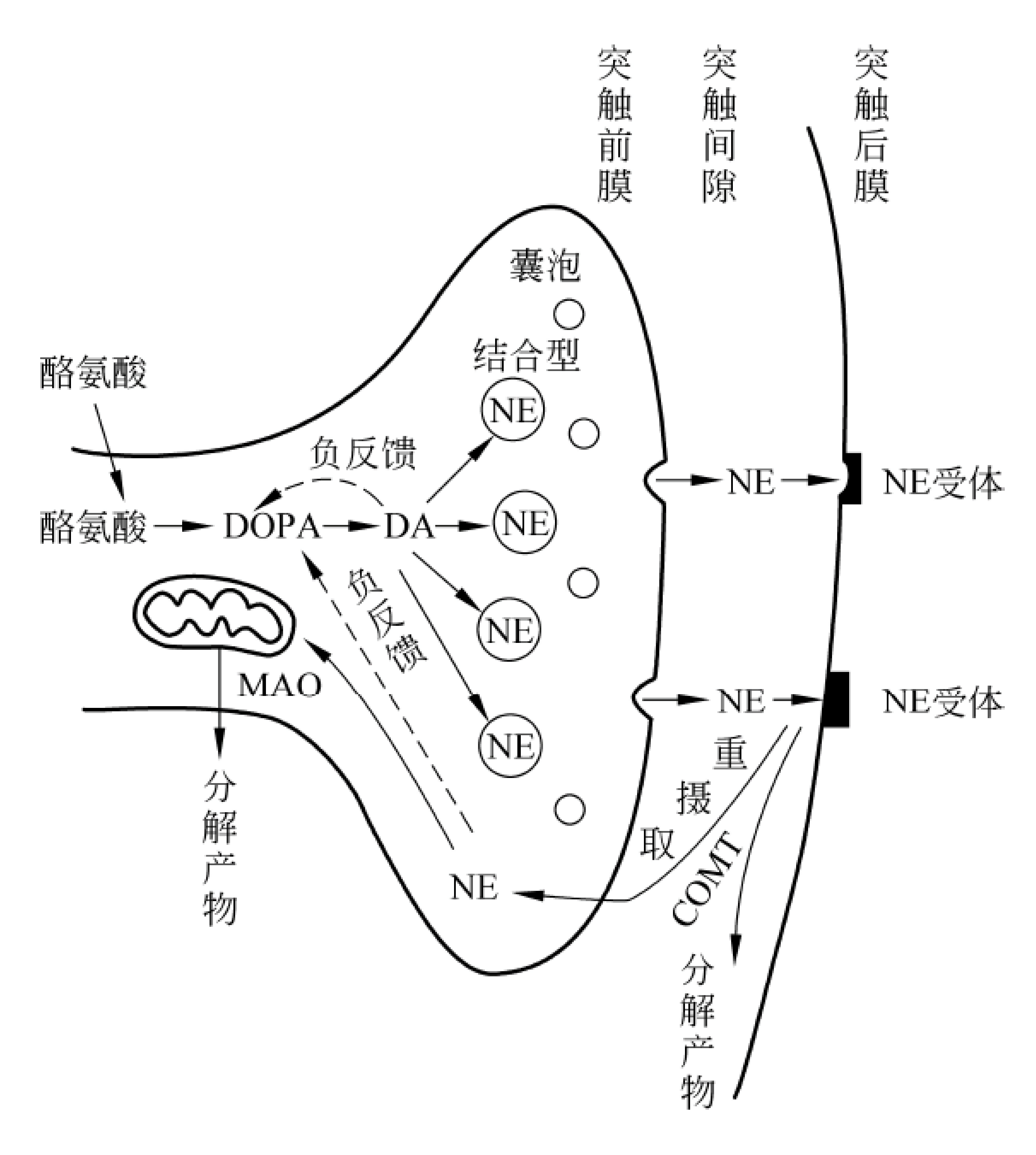



图 2.11　突触部位 NE 的动态

### 3. 多巴胺(DA)

在儿茶酚胺类递质的生物合成过程中,多巴胺是去甲肾上腺素的前体。体内凡有 NE 的组织,其中必然也有多巴胺。由于多巴胺在中枢某些部位中浓度很高,而且它的分布又与去甲肾上腺素不平行,所以一般认为它本身也是一种独立的神经递质。

脑内多巴胺的作用是多方面的,它可能和躯体运动功能的加强,垂体内分泌机能的加强以及精神活动的调节都有关系。

多巴胺能神经末梢中的囊泡是储存多巴胺的场所。这种囊泡不同于去甲肾上腺素囊泡,前者不含多巴胺-β-羟化酶,所以不致将多巴胺羟化成去甲肾上腺素。此外,在去甲肾上腺素能纤维中,去甲肾上腺素囊泡对储存物的要求是 β-位置上有羟基,而多巴胺结构上没有 β-羟基。多巴胺的储存、释放和酶促降解都和去甲肾上腺素十分相似,而更新速度则较去甲肾上腺素快。脑内多巴胺的代谢产物主要是 3-甲氨基-4-羟基苯乙酸(Homovanillic Acid,简写为 HVA,又名同型香草酸):

$CH_3O$—(苯环)—$CH_2$—COOH
HO—

3-甲氨基-4-羟基苯乙酸

## 2.4.3　5-羟色胺

5-羟色胺(5-Hydroxytryptamine,简写为 5-HT)又名血清紧张素(Serotonin),最早是从血清中发现的。中枢神经系统存在着 5-羟色胺能神经元,但在脊椎动物的外周神经系统中至今尚未发现有 5-羟色胺能神经元。

由于 5-羟色胺不能透过血脑屏障,所以中枢的 5-羟色胺是脑内合成的,与外周的 5-羟

色胺不是一个来源。用组织化学的方法证明，5-羟色胺能神经元的胞体在脑内的分布主要集中脑干的中缝核群，其末梢则广泛分布在脑和脊髓中。

5-羟色胺的前体是色氨酸。色氨酸经两步酶促反应，即羟化和脱羧，生成 5-羟色胺，如图 2.12 所示。此过程在某种程度上和儿茶酚胺的生成相似。

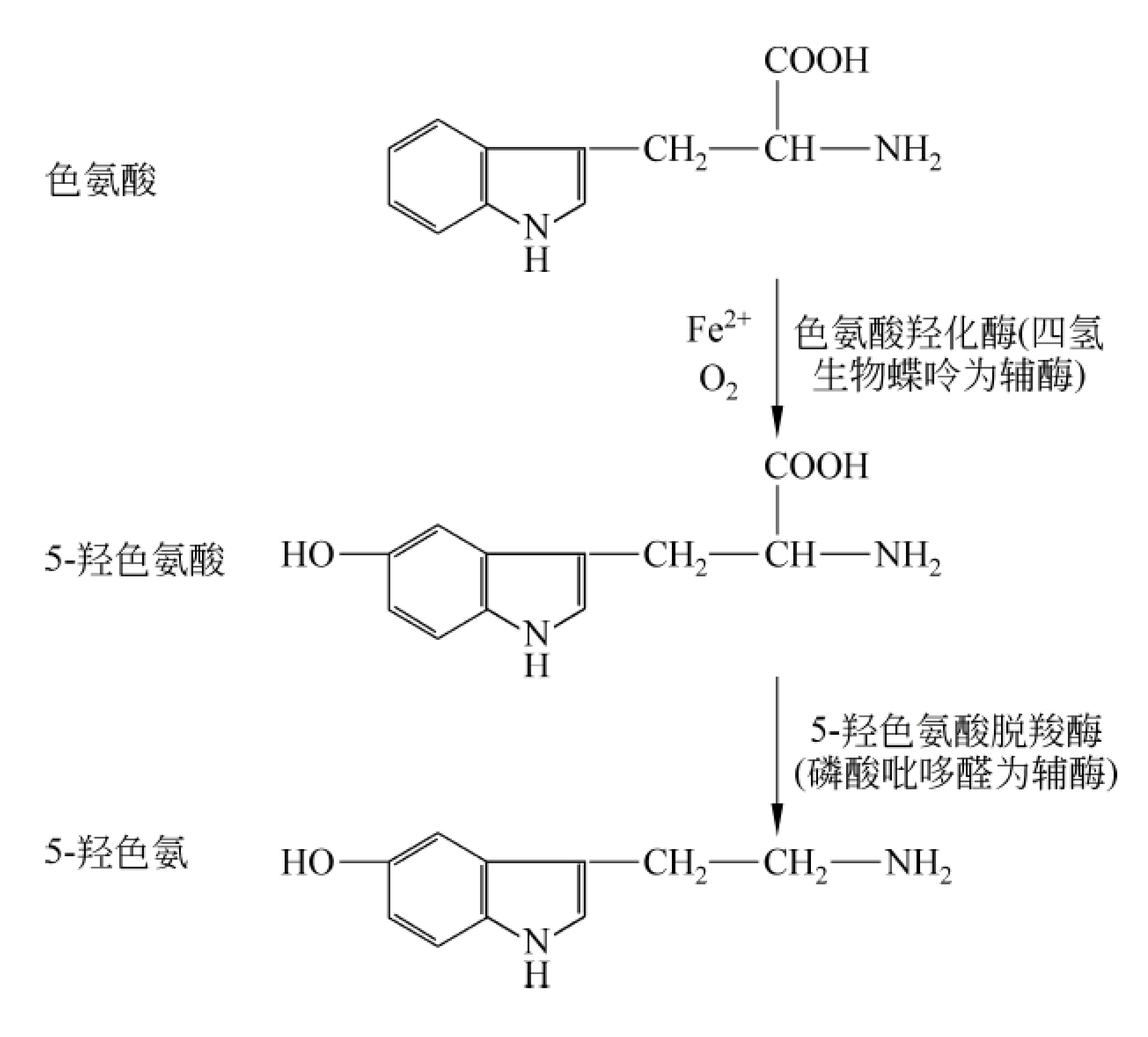



图 2.12　5-HT 的生成

色氨酸羟化酶像酪氨酸羟化酶一样，需要 $O_2$、$Fe^{++}$ 以及辅酶四氢生物蝶呤。但脑内这种酶的含量较少，活性较低，所以它是 5-HT 生物合成的限速酶。此外，脑内 5-HT 的浓度影响色氨酸羟化酶的活性，从而对 5-HT 起着反馈性自我调节作用。血中游离色氨酸的浓度也影响脑内 5-HT 的合成，当血清游离色氨酸增多时(例如给大鼠腹腔注射色氨酸后)，进入脑的色氨酸就增多，从而加速了 5-HT 的合成。

和儿茶酚胺类递质一样，释放到突触间隙的 5-HT，大部分被突触前神经末梢重摄取，而且重摄取后，部分进入囊泡再储存，部分则被线粒体膜上的单胺氧化酶(MAO)所氧化：

HO— —$CH_2$—$CH_2$—$NH_2$ —MAO→ HO— —$CH_2$—COOH

N H　　　　N H

5-羟色胺　　　　5-羟吲哚乙酸

这就是脑内 5-HT 降解的主要方式，5-羟吲哚乙酸(5-Hydroxyindolacetic Acid)无生物活性。

检查 5-HT 对各种神经元的作用时发现，5-HT 可使大多数交感节前神经元兴奋，而使付交感节前神经元抑制。损毁动物的中缝核或用药物阻断 5-HT 合成，都可使脑内 5-HT 含量明显降低，并引起动物睡眠障碍，痛阈降低，同时，吗啡的镇痛作用也减弱或消失。如果电刺激大鼠的中缝核，可影响其体温升高；另外，也观察到室温升高时大鼠脑内 5-HT 更新加速。这些现象揭示脑内 5-HT 与睡眠、镇痛、体温调节都有关系。还有人报道，5-HT 能改变垂体的内分泌机能。此外，有人提出 5-HT 能神经元的破坏是精神性疾病时出现幻觉的原因。可见精神活动也与 5-HT 有一定的关系。

### 2.4.4 氨基酸和寡肽

脑内到处都存在着氨基酸。过去只认为它们是合成蛋白质的原料,或是蛋白质分解的产物。近年来,注意到某些氨基酸在中枢的突触传递中起着递质的作用。而且发现,凡是中性氨基酸,如 γ-氨基丁酸、甘氨酸、β-丙氨酸等对中枢神经元表现抑制作用,而酸性氨基酸如谷氨酸、天门冬氨酸则表现为兴奋作用。

有一些小分子肽类在中枢神经系统中也具有神经递质的作用。1975 年发现的脑啡肽(Enkephalin)是由五个氨基酸残基构成的寡肽,从猪脑中分离出来的脑啡肽又分为两型。

蛋氨酸型(蛋脑啡肽,Met-enkephalin):

H-Tyr-Gly-Gly-Phe-Met-OH

亮氨酸型(亮脑啡肽,Leu-enkephalin):

H-Tyr-Gly-Gly-Phe-Leu-OH

它们是由脑细胞内合成的具有吗啡样作用的肽,故名脑啡肽。已知吗啡(Morphine)要先与脑内的吗啡受体结合才能产生镇痛、欣快作用,而脑啡肽之所以有吗啡样作用也是因为它们能与吗啡受体结合。吗啡是外来的物质,而脑啡肽是内源性的。此外,已知脑内还有一些能与吗啡受体结合并产生吗啡样作用的其他的肽类,称之为内啡肽(Endorphin),虽然还不能十分肯定这些肽类都是真正的神经递质(即完全符合前述神经递质的条件),但是,研究它们的作用对阐明脑的功能,特别对阐明痛觉原理是很有意义的。

### 2.4.5 一氧化氮

在细胞间信息传递的研究领域中出现了一项突破性进展。这个进展是从 1980 年富希宫特(Furchgott)等在家兔大动脉条标本所获得的重要实验结果开始的。他们观察到,当将标本的内皮剥去后,则 Ach 便失去引起该标本平滑肌舒张的作用,然而只要将此无内皮标本与保有内皮的动脉条标本相接触,则 Ach 便可恢复对它的致舒张作用。他们的结论是,Ach 的这种致舒张作用是因为它引起了血管内皮细胞释放出一种被他们称为血管内皮来源的舒张因子的物质。由于此物质极易分解,经数年后方被 Furchgott、Ignarro 和 Murad 三个研究组鉴定为一氧化氮。此三位科学家因为这项研究获得了 1998 年度诺贝尔奖。

一氧化氮为气体,仅微溶于水。这种气体信使物的发现使学术界为之震惊。在一氧化氮的合成酶作用下由 L-精氨酸氧化生成,不经受体的转导,直接扩散通过胞膜作用于靶物质。它的靶物质很多,其中有代表性者为触媒 GTP 生成 cGMP 的鸟甘酸环化酶。此酶被一氧化氮激活,通过转导通路降低胞内 $Ca^2$,导致平滑肌舒张。

在神经型一氧化氮合成酶作用下生成的一氧化氮作为信使物参与中枢神经系统中突触可塑性、记忆、视觉和嗅觉等功能。在周边神经系统中主要是非肾上腺素能和非胆碱能神经元的信使物。

### 2.4.6 受体

受体(Receptor)的概念最早是由英国药理学家兰格列(A D Langley)于 1905 年提出的。他观察到箭毒虽可拮抗烟碱引起的肌肉收缩,但不能阻遏直接电刺激肌肉引起的收缩,

由此他设想这两种试剂都只和细胞中非神经性和非肌肉性的特定物质相结合，所不同的只是烟碱与该特定物质结合可进一步产生生物效应，即肌肉收缩，但箭毒结合所产生的效应则是拮抗烟碱引起的肌肉收缩。他把该特定物质称为接受物质。事实上，兰格列当时提出的关于接受物质（受体）的概念现在看来仍是正确的，因为他已指出了受体的两个重要特征，即识别特定物质和产生生物效应。

烟碱和箭毒等可选择地作用于体内特定分子并能引起生物效应的物质称生物活性物质。如体内固有的，则称为内源性活性物质，如递质、激素、营养因子等；如来自体外者，则称外源性活性物质，如烟碱、箭毒等药物和毒物。那些在胞膜以及胞浆与核中对特定生物活性物质具有识别并与之结合而产生生物效应的大分子被称为受体，而那些与受体有选择性结合特性的生物活性物质称配体，其中与受体结合可引起生物效应者称为激动物（Agonist），与受体结合，但其生物效应表现为选择地拮抗由激动物引起的生物效应者称拮抗物。至于既有激动作用又有拮抗作用者称为部分激动物。神经元上的受体称神经受体。

自从 1983 年奴曼（Numa）和同事成功地纯化并测定了 AchR 的一级结构，从而开辟了研究受体分子结构与功能的先河以来，已有众多的受体及其亚型的一级结构被确定，使得我们有可能按分子结构特征对受体进行分类，并对受体分子的活动机制进行分析。

一般认为，用生化方法鉴定受体应具备下列 3 个特性：①饱和性。受体分子的数量是有限的，因此配体与受体结合的剂量效应曲线应具饱和性，并且它们的特异结合应表现为高亲和性和低容量性。细胞往往对配体也可能有非特异性结合，但这种结合既表现为低亲和性和高容量性，又无饱和性。②特异性。特定的受体只与特定配体结合产生生物效应。因此常用比较一系列配体的生物效应的方法来研究受体特性，并对之进行功能分类。③可逆性。在生理活动中配体和受体的结合应是可逆的。配体与受体复合物的解离常数虽有不同，但被解离下来的配体应是原物，而不是其代谢产物。

受体的两个主要功能是选择地识别递质和激活效应器，因此便可按所选择识别的递质将它们分为 AchR、GluR、GABA 受体、GlyR、5-HTR、组胺受体和识别 NA 和 Ad 的 Ad 受体，以及识别各种神经肽的受体等。另一方面，由于已阐明了多种受体分子的一级结构，故又可按它们作用于效应器的分子机制将其分为直接调控离子通道活动的离子通道型受体和间接调控离子通道活动的代谢调节型受体。离子通道型受体分子中既有识别递质的受点，又有离子通道，故活动速度快。已知属此类型的有：nAchR、$GABA_A$R（A 型 GABA 受体）、5-$HT_3$R 和 iGluR（离子通道型谷氨酸受体）。代谢调节型受体分子中只含识别递质的位点，并无容许离子通过的微孔道。它们与效应器的功能耦联是经鸟嘌呤核苷酸结合蛋白，即G-蛋白实现的，因此这类受体又被称为 G 蛋白耦联受体。已知属此类型者又可分为两个基因家族，包括 mAchR、$GABA_B$R、mGluR（代谢调节型谷氨酸受体）和 5-$HT_{1,2,4}$R（1，2，4 型 5-HTR），以及各种神经肽受体等。此外尚有由营养因子、激素和某些神经肽激活的、穿膜一次并在胞内 C 末端带有（或不带有）激酶的受体。

## 2.5　信号跨膜转导

生命信号转导简言之就是：外环境刺激因子和机体、胞间通信信号分子——激素、神经递质等第一信使、配体，到达并作用于机体细胞表面或胞内受体部位后，跨膜信号转换，形成

胞内第二信使,以及经过其后的信息、信号途径组份级联传递(蛋白质可逆磷酸化传递信息),通过胞内信号将信息传递到特定效应部位而起作用,引起细胞生理反应和诱导基因表达的过程。其经典转导途径简示如下:

外、内刺激→配体→受体→跨膜→第二信使→磷酸化传递信息→调节细胞机体

水溶性神经递质一般都不进人胞内,只与胞膜上的受体结合,将信号传给受体便完成了使命。离子通道型受体接受传来的信号便直接通过藏于自身分子内的效应器,即离子通道给出反应,但代谢调节型受体则不同,它们必须经一系列的转传过程将信号传入脑内,再传给离子通道或代谢型效应器分子。受体将接受的信号跨膜转传给离子通道或胞内的代谢型效应器分子的过程称信号(跨膜)转导。近二十年来,这一领域的研究有了快速发展。这一快速发展进程是从激素作用的第二信号学说的提出和其后的转导蛋白(G蛋白)的发现开始的。在激素作用机制的研究中提出的学说和新发现又推动了神经递质和其他生物活性物质作用机制的研究。

## 2.5.1 转导蛋白

1958年苏瑟兰德(E W Sutherland)发现,胰高血糖素和肾上腺素促使肝细胞糖原分解作用都是通过一种由他发现的新物质,即cAMP(Cyclic Adenosine Monophosphate,环单磷酸腺苷)实现的。这两种激素与各自的受体结合都激活位于膜中的腺甘酸环化酶(Acase),由后者催化ATP生成cAMP。他还发现cAMP是激活cAMP依赖蛋白激酶(cAMP-dependent Protein kinase,PAK)所必需,由活化的PAK再激活磷酸酯酶从而催化肝糖原的分解。在这些研究结果的基础上,苏瑟兰德于1965年便提出了激素作用的第二信使理论,即激素不进入脑内,只是作为第一信使将信号带给膜上的特定受体,再由受体激活膜中的ACase系统,将胞内的ATP转化为cAMP。另一方面,cAMP作为第二信使将信号送到效应器分子给出反应。由于这一学说的提出又促成了转导蛋白,即G蛋白(Guanosine Nucleotide Binding Protein)的发现。

洛德贝尔(M Rodbell)在脂肪细胞碎片制备的实验中证明了几种激素(如胰高血糖素、Ad和ACTH等)都通过ACase系统催化ATP生成cAMP这一共同的第二信使。于1970年他又发现,在由激素、受体、ACase系统催化ATP转化为cAMP的反应中尚必须有GTP的参与。实验又表明,GTP在反应中不被分解,又不与受体和ACase结合。最后,由基勒曼(A G Gilman)于1987年成功地将这种与GTP结合,又可将受体与ACase联系起来的转导蛋白从膜制备中分离纯化了出来,称GTP结合蛋白。这是最早提出的信号跨膜转导途径。为此苏瑟兰德获得了1971年度以及基勒曼和洛德贝尔获得了1994年度的诺贝尔奖。

参与神经元的信号跨膜传导的G蛋白是一组分子量约10万的可溶性膜内蛋白,按功能作用分为:Gs、Gi、Go和Gt 4种。Gs为激活ACase活性者,其药理特性是可被霍乱毒素(CTX)直接激活,它可开启某些Ca通道。Gi为抑制ACase活性者,对百日咳毒素(PTX)敏感,受其作用后便不再能被受体激活。它可开启某些K通道,给出抑制性反应。Go原初是指不作用于ACase的其他G蛋白,但现在则仅代表其中对PTX敏感,激活PKC的G-蛋白,而将其中激活PKC,但对PTX不敏感者称为Gpc型,Go和Gpc富含于脑组织中,可开启某些K通道,也可以抑制T和N型电压门控Ca通道。Gt是耦联光感受器中视紫质和被级效应器酶cGMP-PDE的G蛋白,它可使Na通道关闭。

受体在静息时处于与配体有高亲和力的活化状态，G-GDP 为非活化型。当配体（A）作用于受体（R）而形成 A·R，因 A·R 与 G-GDP 的亲和性高，于是便形成 A·R·G-GDP 复合体。此复合体的形成引起 GDP 被胞浆中 GTP 置换的反应，形成活化型 G-GTP。此过程是在受体与配体结合而失活的同时，G 蛋白由非活化型转成活化型。在复合体完成 GTP 置换的同时便分解成 A·R＋α-GTP＋βγ 三部分。其中 α-GTP 具有水解 GTP 成为 GDP 的酶活性（催化速率低），因而在 α-GTP 激活效应器（E）的同时，失去磷酸分子而变成 α-GDP。α-GDP 与 βγ 亚基亲和性高而再合成为 G-GDP，即恢复到静息时的非活化型。另外，A·R 由于和 G-GDP 的结合，便降低了与 A 的亲和性，使之脱离，于是 R 也恢复到静息时的活化状态。此过程又是在使 G-蛋白激活效应器而转化为非活化型的同时，受体则由非活化型恢复为活化型。

## 2.5.2　第二信使

作为第一信使的递质及内源性活性物质已有近百种，它们的受体总数可达 2～3 倍之多，其中多数是以某种形式依靠 G 蛋白与效应器耦联的，即代谢调节型受体。已知膜磷脂水解而生成的第二信使有：三磷酸肌醇（Inositol Triphosphate，$IP_3$）和甘油二酯（Diacyl Glycerol，DG）以及花生四烯酸（Aracidonic Acid）等，信号的跨膜转导也相应地被分为：cAMP、磷酸肌醇和花生四烯酸系统。近年来又发现作为第二信使的尚有一氧化氮合成酶催化生成的 NO，还可能有 CO。

第二信使通路有共同的转导方式。如图 2.13[967] 所示，3 个系统的活动方式都可分为 3 个步骤：从胞外的第一信使，经膜上的受体、结合在膜内侧的转导蛋白和膜中的初级效应器（酶），到胞内的第二信使和次级效应器。不同的递质分别作用于各自的受体，通过某种 G 蛋白的转导，激活各自的初级效应器，传至相应的第二信使，再激活次级效应器等。

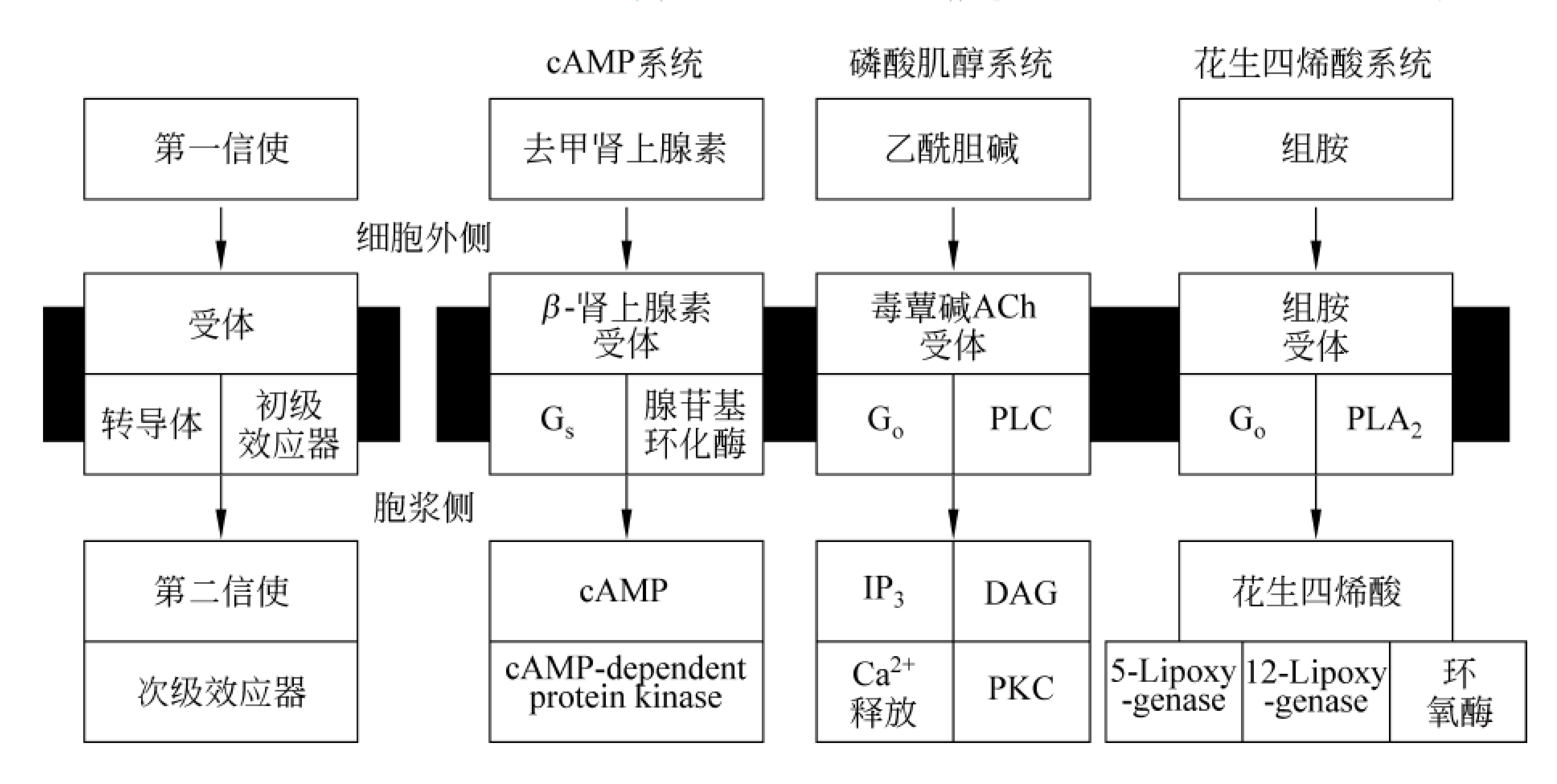



图 2.13　信号跨膜转导通路示意图

多数次级效应器为激酶。由于一种激酶往往可磷酸化不同的靶蛋白，所以各系统往往可能导致信号的放大和交叉，最终引起离子通道的开放或关闭，或发生相应的代谢变化。又由于 G 蛋白分子的数量远大于受体分子的数量，因此 1 个与递质结合了的受体可激活多个 G 蛋白分子，从而可将转导的信号放大。

已知在cAMP系统中的次级效应器酶,环化腺苷酯依赖的蛋白激酶(cAMP-dependent Protein Kinase)分子由2个调节亚基和2个催化亚基(C)构成。其中2个相同的R亚基都有4个区段:①N末端区段为与同伴亚基间的结合位点;②接着的区段与C亚基结合抑制其酶活性;③2个结合cAMP的相同区段。整个PKA分子可写成$R_2C_2$。当在ACase作用下,ATP转化为cAMP时,则

$$R_2C_2+4cAMP=2(R\cdot 2cAMP)+2C$$

由此生成的C便可进一步使效应器蛋白磷酸化(见图2.14)。

在磷酸肌醇系统中,如图2.13所示,首先由PLC比水解磷酸肌醇生成$IP_3$和DAG。DAG为疏水性而留在胞膜内以激活PKC。但非活化的PKC在胞浆中。当生成DG时,PKC便去与膜中的DG结合而被激活,于是可使膜内或胞浆内的底物蛋白磷酸化,另一个第二信使为$IP_3$。某些受体导致$IP_3$的生成,可使作为胞内$Ca^{2+}$库的肌浆网等膜上的Ca通道开放,升高胞内$Ca^{2+}$浓度,再由$Ca^{2+}$激发各细胞活动。

在花生四烯酸系统中是由$PLA_2$,催化从膜中释放出花生四烯酸,它生成后立即被如图2.13所示的3个酶转化成几个活性代谢产物。

除上述3个第二信使系统之外,尚有首先在视杆和视锥细胞中发现的cGMP系统。光作用于视紫质,经Gt的转导而生成的第二信使cGMP,可直接调控Na通道。在黑暗中cGMP浓度较高,可开启cGMP门控Na通道,因而导致视感觉器细胞的相对去极化。当光激活视紫质,后者刺激cGMP磷酸二酯酶,降低cGMP浓度,从而关闭通道,导致光感受器细胞发生超级化。

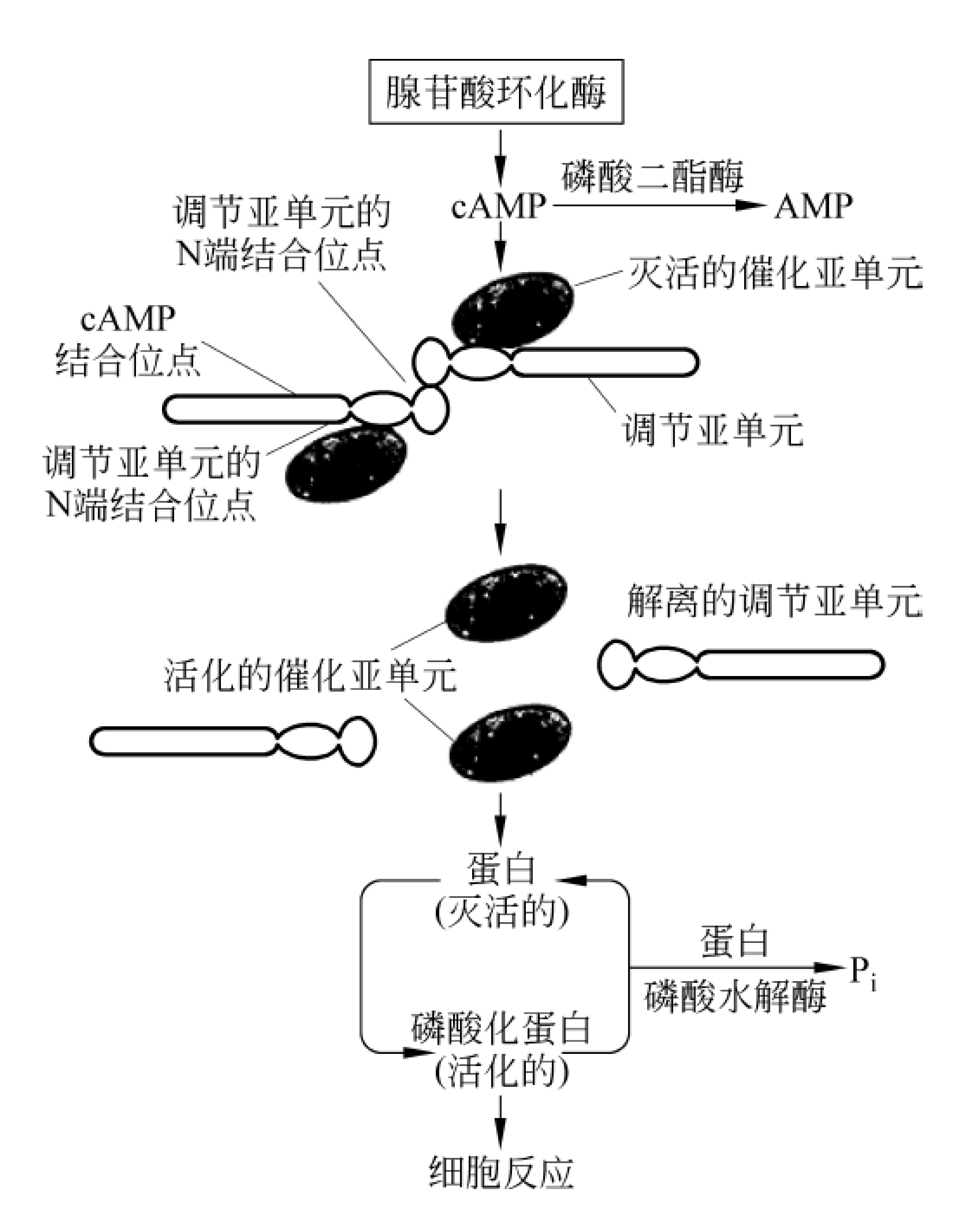



图2.14　cAMP信号转导通路中的分子机制

由上可知,从G蛋白调控离子通道角度看,大致可有几种方式:①G蛋白的α或βγ亚基直接作用于离子通道,至少在K和Ca通道上发现了这种调控方式;②G蛋白通过效应

器酶和第二信使再作用于离子通道；③G 蛋白通过效应器酶和第二信使，再经蛋白激酶作用于通道。

近年来有资料表明，神经细胞的膜和核受体在外来因子，如递质、调质、营养因子和激素等的作用下，通过 G 蛋白的转导和胞内效应器的磷酸化，将信号传入核内，诱发基因表达的变化。这种经第一与第二信使的转导，诱发基因表达变化，与突触可塑性和学习与记忆机制有密切关系，其中被研究较多的是直接早期基因和延迟早期基因。

中科院上海生命科学研究院神经科学研究所周专研究员及其学生张晨首次发现在老鼠的一种感觉背根神经节细胞上，神经递质不仅可由钙离子指挥而释放到下一级神经细胞，而且电压冲动本身也可以完全独立地导致神经递质的释放。这种与钙离子无关的单纯由神经电冲动导致的神经信号传导机制是一个相当出乎意料的新发现，向神经科学研究提出了一系列新课题。

神经信号是如何从一个神经细胞传到另一个神经细胞的？这是神经科学家们研究的一个焦点问题。当前占统治地位的学术观点是：前一级神经细胞在“兴奋”时将产生一个电冲动（“动作电位”），在电冲动期间细胞外的钙离子可以流入该细胞内，流入的钙离子导致该细胞分泌一些活性分子（神经递质）传递到相邻的下一级神经细胞的外表面（细胞膜）。细胞膜表面的“受体”膜蛋白分子与神经递质结合，即可导致第二级神经细胞产生电冲动。以此类推，神经信号便一级一级地传递下去，从而构成复杂的信号体系，乃至最终出现学习、记忆等大脑的高级功能。这就是钙离子指挥的神经递质释放和神经信号转导机制。周专研究员及其学生张晨发现，与钙离子无关的、单纯由神经电压冲动导致的神经信号转导，可能对神经信号的转导和信息整合研究产生重要影响。这一发现还向神经科学研究提出一系列新问题。据周专透露，至少在某些神经类细胞（例如肾上腺嗜铬细胞上）只存在钙离子导致的神经递质分泌途径。此外，这种“非钙电压—分泌耦联”信号转导的分子机制也成为神经科学研究需要进一步解决的重要课题。

## 2.6　静息膜电位

生物电是在研究神经与肌肉活动中首先被发现的。意大利医生和生理学家伽尔凡尼（L Galvani）在 18 世纪末进行的所谓“凉台实验”是生物电研究的开端。当他把剥去皮肤的蛙下肢标本用铜钩挂到凉台的铁栏杆上，以便观察闪电对神经肌肉的作用时，意外地发现每当蛙腿肌肉被风吹动而触及铁栏杆便出现收缩。伽尔凡尼认为，这是生物电存在的证明。

1827 年物理学家依贝利（Nobeli）改进了电流计，并在肌肉的横切面和完整的纵表面之间记到了电流，其损伤处为负，完整部分为正。这是首次实现了对生物电（损伤电位）的直接测量。德国生理学家雷蒙德（D B Reymond）一方面改进和设计了许多研究生物电现象的设备和仪器，如电键、乏极化电极、感应线圈和更为灵敏的电流计等；另一方面，又对生物电进行了广泛和深入的研究，如在大脑皮质、腺体、皮肤和眼球等生物组织或器官都发现了生物电，特别是 1849 他又在神经干上记录到损伤电位和活动时产生的负电变化，即神经的静息电位和动作电位，并且在此基础上首次提出了关于生物电产生机制的学说，即极化分子说。他设想神经肌肉细胞表面是由排列整齐的、宛如磁体的极化分子构成。每个分子的中央有一条正电荷带，两侧均带负电荷。正电荷汇集于神经与肌肉的纵表面，在它们的横断面上汇集的便是负电荷。因此，在神经与肌肉表面和内部之间形成了电位差。当神经与肌肉兴奋

时,它们的排列整齐的极化分子变为无序状态,表面与内部的电位差消失。

雷蒙德的一位学生勃斯特恩(Bernstein)在电化学进展的影响下,发展了生物电的既存说,提出了现在看来仍相当正确的,并推动了生物电研究的膜学说。这一学说认为,电位存在于神经和肌细胞膜的两侧。在静息状态,胞膜只对 $K^+$ 有通透性,对较大的正离子和负离子均无通透性。由于膜对 $K^+$ 的选择性通透和膜内外存在的 $K^+$ 浓度差,便产生了静息电位。当神经兴奋时,胞膜对 $K^+$ 的这种选择性通透的瞬时丧失变成无选择性通透,导致膜两侧电位差的瞬时消失,便形成了动作电位。

20 世纪 20 年代,伽塞(H S Gasser)和厄兰格(J Erlanger)将阴极射线示波器等近代电子学设备引入神经生理学研究,促进了生物电研究的较快发展。于 1944 年他们两位由于对神经纤维电活动的分析而共同获得了诺贝尔奖。杨(Young)报道了乌贼神经干中含有直径达 500$\mu$m 的巨轴突。英国生理学家霍奇金(A L Hodgkin)和胡克列(A F Huxley)将毛细玻璃管电极从切口纵向插入该巨轴突内首次实现了静息电位和动作电位的胞内记录,并在对这两种电位的精确定量分析的基础上,证实并发展了勃斯特恩关于静息电位膜学说的同时,又提出了动作电位的钠学说。接着他们又进一步应用电压钳技术在乌贼巨轴突上记录了动作电流,并证明它可被分成 Na 与 K 电流两个成分。在此研究的基础上他们又提出了双离子通道模型,指引了离子通道分子生物学的研究。在微电极记录技术的推动下,神经细胞生理学的研究又步入了新的发展时期。埃克勒斯(S J Eccles)开始应用玻璃微电极对脊髓神经元及其突触的在位的电生理研究,发现了兴奋性和抑制性突触后电位。基于对神经生理学研究的贡献,霍奇金、胡克列和埃克勒斯三人分享了 1963 年的生理学或医学诺贝尔奖。珈兹(S B Katz)则开始应用微电极技术开展了神经肌肉接头突触的研究,为此于 1970 年也获得了诺贝尔奖。在神经系统研究的蓬勃发展的基础上,于 20 世纪 60 年代便形成了神经系统研究的综合学科,即神经生物学和神经科学。

静息膜电位是神经与肌肉等可兴奋细胞的最基本的电现象,因为当它们活动时所发生的各类瞬时电变化,如感受器电位、突触电位和动作电位等都是在此静息膜电位的基础上所发生的瞬时变化。为了描述方便通常把胞膜两侧存在电位差的状态称为极化,并且将静息膜电位绝对值向增加方向的变化称为超极化,以及向减少方向的变化称为去极化(如图 2.15 所示)。

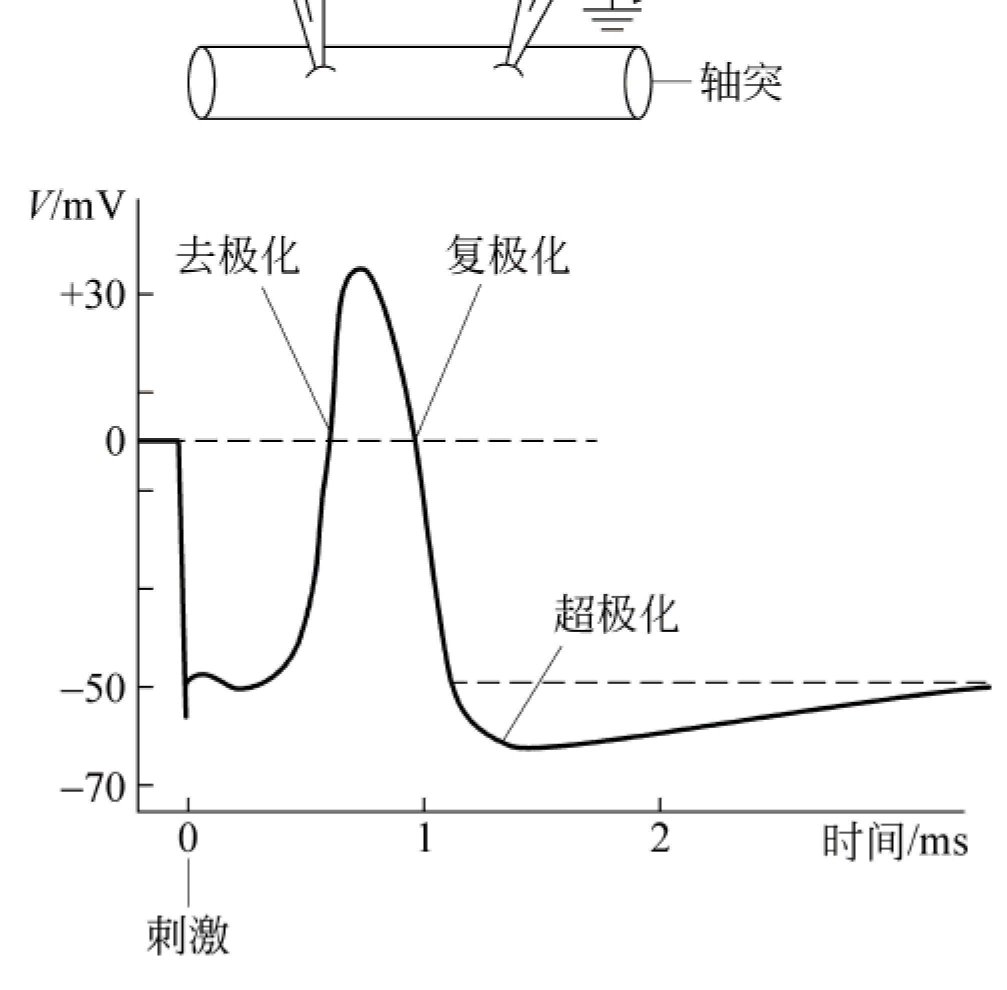



图 2.15　胞膜两侧存在的电位差

在处于静息状态的神经和肌肉等可兴奋细胞膜的两侧存在着高达约 70mV 的电位差。这提示在它们的胞膜的内侧面与外表面分别有负与正的离子云的分布，即分别有多余的负与正离子的汇聚。在神经元胞浆内所含离子中可以说没有一种，其浓度与胞外体液中的是相同的，特别是其中的 K、Na 和 Cl 三种离子，不但其胞内与胞外的浓度均达 mmol/L 水平（称常量离子），并且跨膜浓度差又均约为 1 个数量级。$Na^+$ 与 $Cl^-$ 富集于胞外，而 $K^+$ 则富集于胞内。还有一些大的有机负离子（$A^-$）可以认为只含于胞内，其总浓度也在 mmol/L 水平。

由连续的类脂双层构成的胞膜中分散地镶嵌着被称为离子通道的大蛋白分子。它们横贯胞膜，在其分子中轴含有亲水性微孔道，可选择地容许特定离子通过。按它们可通过的离子种类，如 $K^+$、$Na^+$、$Cl^-$ 和 $Ca^{2+}$，而分别被称为 K、Na、Cl 和 Ca 通道。离子通道至少有两种状态，即开放态和关闭态。离子通道开放便会有特定离子顺浓度差跨膜移动。静息膜电位便是由可兴奋细胞膜中的在静息状态持续开放的所谓静息离子通道容许特定离子沿其浓度梯度跨膜移动而形成的。

神经元胞膜对电流为起着电阻作用。这种电阻称膜电阻。除电阻作用外，胞膜尚起着电容器作用。这种电容称为膜电容。可以采用图 2.16 的连接，测量膜电位的变化。当向胞膜通电流或断去电流时都要分别地先经电容器的充电或放电过程，从而使得电紧张电位的上升和下降时均以指数曲线变化。如在 $t=0$ 将电流注射入胞内，经任意时间 $t$ 所记录到的电位称为 $V_t$，则

$$V_t = V_\infty(1 - e^{-t/\tau}) \tag{2.1}$$

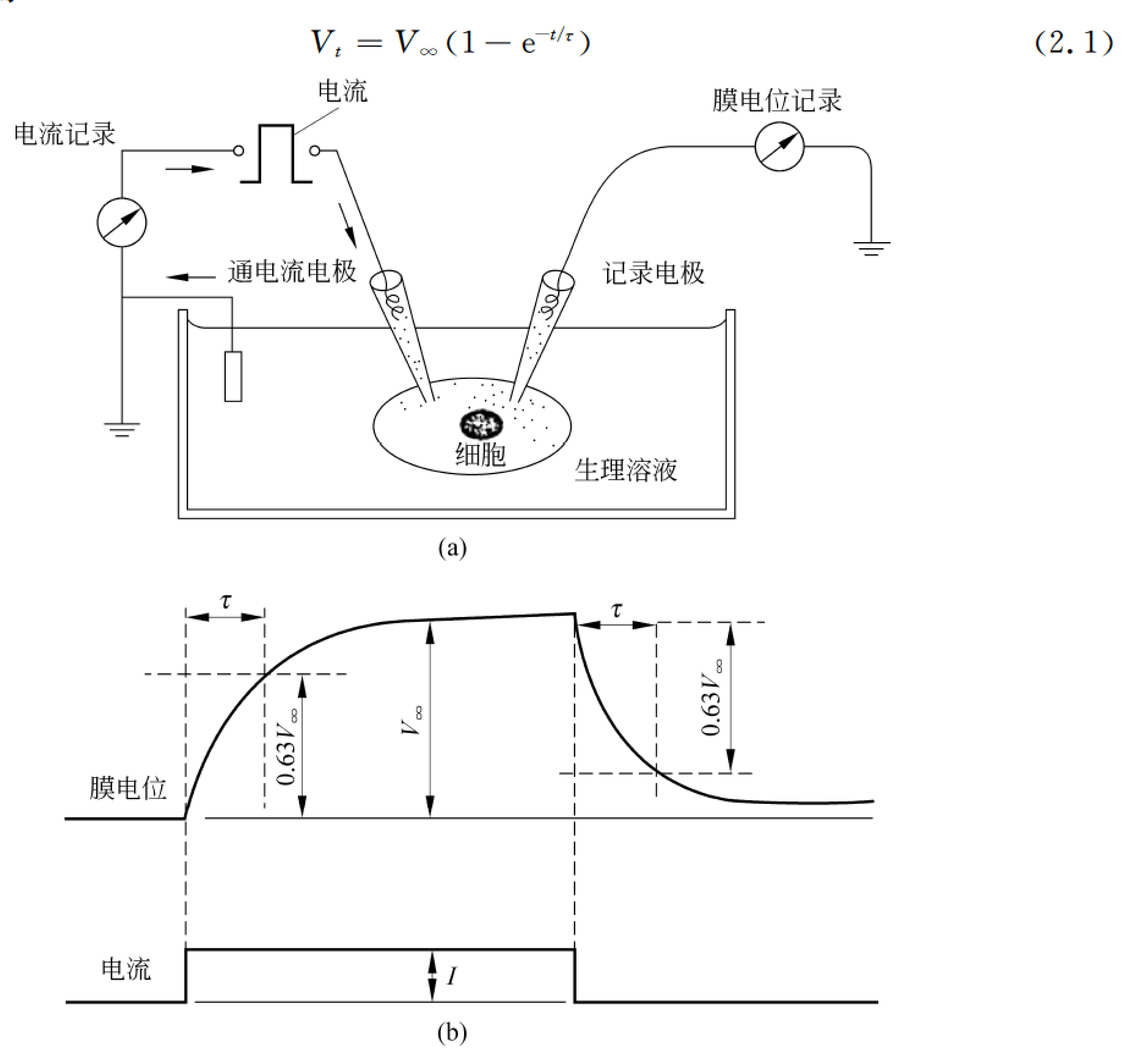



图 2.16　神经元胞膜电位

式中 $V_\infty$ 为电容充电完成后的恒定电位值。不难看出,当 $t=\tau$ 时,式(2.1)可简化为

$$V_t = V_\infty\left(1-\frac{1}{\mathrm{e}}\right) = 0.63V_\infty \tag{2.2}$$

即 $\tau$ 为电紧张电压升至 $0.63V_\infty$ 时所需时间。于是,就把 $\tau$ 定为表示膜的电紧张电位的变化速度的时间常数,它应等于膜电容 $C$ 与膜电阻 $R$ 的乘积,即

$$\tau = RC$$

其中 $R$ 可在实验中用通电电流值去除 $V_\infty$ 的值求得。这样便可分别测出膜电阻和求出膜电容值。为了对各种可兴奋细胞膜的电学性质进行比较,通常还进一步求出膜单位面积的比膜电阻和比膜电容值。膜电容来自膜的类脂双层,膜电阻来自膜中的离子通道。

## 2.7 动作电位

神经元具有两种基本特性:兴奋和传导。当神经元的某一部分受到某种刺激时,在受刺激的部位就产生兴奋。这种兴奋会沿着神经元散布开来,并在适当的条件下通过突触传达到与之相联的神经细胞,或传达到其他细胞,从而使最后传达到的器官的活动或状态发生变化。细胞受刺激后在静息电位基础上发生的一次膜两侧电位快速倒转和复原,称为动作电位(见图2.15)。一定强度的阈下刺激所诱发的局部电位是随刺激的增强而变大,但动作电位则不同,在阈下刺激时根本不出现。当刺激一旦达到阈值以及超过阈值,便在局部电位的基础上出现,并且自我再生地快速达到固定的最大值,旋又恢复到原初的静息膜电位水平。这种反应方式称全或无反应。

动作电位另一个特性是不衰减传导。动作电位作为电脉冲,它一旦在神经元的一处发生,则该处的膜电位便爆发式变为内正外负,于是该处便成为电池,对仍处于静息膜电位(内负外正)的相邻部位形成刺激,并且其强度明显超过阈值。因此相邻部位因受到阈上刺激而进入兴奋状态,并且也随之产生全或无式动作电位。这样,在神经元一处产生的动作电位便以这种局部电流机制依次诱发相邻部位产生动作电位,又由于动作电位是全或无式反应,所以它可不衰减地向远距离传导。但在轴突末梢,因其直径变小而动作电位振幅亦随之变小。

在神经元膜的某处一旦发生了动作电位,则该处的兴奋性便将发生一系列变化。大致在动作电位的超射时相,无论用如何强的刺激电流在该处都不能引起动作电位,此时相称为绝对不应期;在随后的短时间内,用较强的闭上刺激方可以在该处引起动作电位,并且其振幅还要小一些,此时相称为相对不应期。如动作电位的持续时间为1ms,则这两时相加到一起应不超过1ms,因为否则前后两个动作电位将发生融合。

动作电位主要生理功能如下。

(1) 作为快速而长距离地传导的电信号。

(2) 调控神经递质的释放、肌肉的收缩和腺体的分泌等。

各种可兴奋细胞的动作电位虽有共同性,但它们的振幅、形状,甚至产生的离子基础也有一定程度的差异。

动作电位超射时相的发现便否定了勃斯特恩经典膜学说对动作电位的解释,即将动作电位的发生归于胞膜对离子选择性通透的瞬间消失的观点是不能成立的。在20世纪50年代初霍奇金等在乌贼巨轴突上进行的精确实验表明,静息状态时轴突膜的 $K^+$、$Na^+$ 和 $Cl^-$

通透系数为：$P_K$ ：$P_{Na}$ ：$P_{Cl}$ = 1 ：0.04 ：0.45，在动作电位顶峰时这些系数比变为：$P_K$ ：$P_{Na}$ ：$P_{Cl}$ = 1 ：20 ：0.45。很显然，$P_K$ 与 $P_{Cl}$ 的比例未变，只是 $P_K$ 与 $P_{Na}$ 之比显著增大了 3 个数量级。根据这些及其他一些实验资料，他们便提出了动作电位的钠离子学说，即认为动作电位的发生取决于胞膜的 $Na^+$ 通透性的瞬时升高。换句话说，动作电位的发生是胞膜从主要以 $K^+$ 平衡电位为主的静息状态突变到主要以 $Na^+$ 平衡电位为主的活动状态。

图 2.17 给出了 Na 与 K 电导在动作电位过程的变化[324]。从图 2.17 中可以看出，动作电位的发生乃是轴突膜的 $Na^+$ 通透性的迅速升高，使膜电位接近 Na 平衡电位，随后又迅速下降，随之又有 $K^+$ 通透性的持续上升，使膜电位恢复到静息时的接近 $K^+$ 平衡电位水平。Na 电导的下降有两种不同的方式：将膜电位从静息水平固定在 −9mV，再在图 2.17 上图中左侧虚线处，即在短时间内（此例为 0.63ms）使膜电位恢复原初水平，则 Na 电导也随之快速消失（图 2.17 中图中虚线）。此时若再度使膜电位去极化，则 Na 电导仍可出现；另一种情况是，如膜电位被跃迁至 −9mV 后，一直持续在该水平，则 Na 电导也会逐渐变小，直至消失。若此时再度使膜电位去极化，则不会有 Na 电导出现。此现象被称为 Na 电流的灭活。这时必须在膜电位恢复数毫秒后，第二个刺激方可能成为有效的，这种从灭活恢复的过程称去灭活。K 电导则有所不同，在去极化持续约 6ms 多（图 2.17 上边中右侧虚线处）仍维持在最高水平，并且在膜电位恢复原初水平时，以与出现时相反的曲线消失。由于 K 电流的去活化很慢，曾被认为无去活化过程。

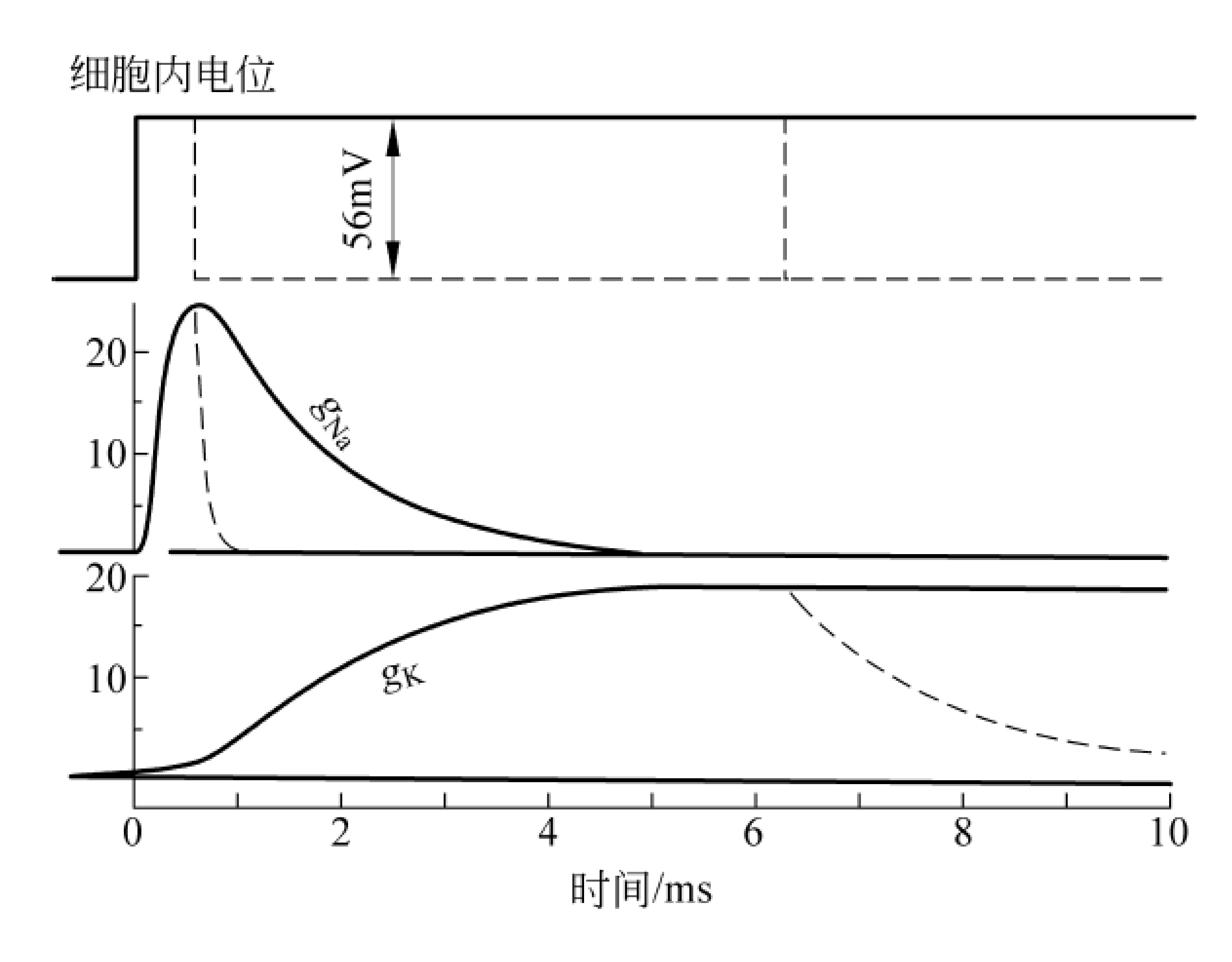



图 2.17　乌贼巨轴突去极化 56mV 时 Na 与 K 电导的变化

乌贼巨轴突的动作电流是由内向 $Na^+$ 流和迟出的外向 $K^+$ 流合成的，而这两股离子流又是两种离子分别通过各自的电压门控通道进行跨膜流动而产生的。在乌贼巨轴突上取得了进展之后，关于用电压钳技术分析动作电流的研究便迅速地被扩大到其他可兴奋细胞。结果发现这两种电压门控通道几乎存在于所有被研究过的可兴奋细胞膜中，此外又发现了电压门控 Ca 通道。在某些神经元还发现有电压门控 Cl 通道。这 4 种电压门控通道又有不同类型，至少 Na 通道有两种类型（在神经元发现的神经型和在肌肉发现的肌肉型）、Ca 通道有 4 种类型（T，N，L 和 P 型），K 通道主要有 4 种类型（延搁整流 K 通道、快瞬时 K 通道或称 A 通道、异常整流 K 通道和 $Ca^{2+}$ 激活 K 通道）。产生动作电位的细胞称可兴奋细胞，但不同类型的可兴奋细胞产生的动作电位的振幅和时程有所不同。这是因为参与形成这些

动作电位的离子通道的类型和数量的不同。

轴突膜和多种神经元胞体膜的动作电位的形成机制较为单纯,其上升相都是由 $Na^+$ 流形成的,称 Na 依赖动作电位。这类动作电位的振幅较大,持续时间也较短,其传导速度也比较快。至于在轴突动作电位下降相中有否 K 电流成分的参与,则依动物种类而有所不同,如在家兔有髓神经纤维郎氏结膜的动作电位中便与乌贼巨轴突中的不同,无 K 电流成分参与,至于在青蛙郎氏结轴突膜,特别是无脊椎动物对虾的兴奋结轴突膜动作电位中虽有 K 电流成分,但它不但很小,并且激活阈也较高,因此对缩短动作电位持续时间应无明显作用。

不同类型的神经元有不同的兴奋性,即使一个神经元的不同部位,如轴丘、轴突末梢和树突等处的兴奋性也有差异。兴奋性的不同便是由兴奋膜中的电压门控通道的种类和密度决定的。有些神经元胞体和轴突末梢的动作电位是由 Na 与 Ca 电流共同形成的,这种动作电位的持续时间比较长。在某些神经元的树突还发现由 Ca 电流形成的动作电位,其振幅小和持续时间长。

动作电位,即神经冲动一旦在神经元(细的树突除外)一处产生,便以恒定的速度和振幅传到其余部分。动作电位爆发式出现时,即由局部电位发展为动作电位时的膜电位称阈电位。从静息膜电位至阈电位的去极化通常称为临界去极化。此临界去极化约为 32mV。一般说来,阈值是和静息膜电位与阈电位之差,即临界去极化成正比,所以阈值随这两个电位的相对变化而变化。所谓阈电位乃是由去极化引起的 $Na^+$ 通透性的升高达到 $Na^+$ 的内流量恰好与 $K^+$ 的外流量相等时的膜电位。至于局部电位的出现乃是 $g_{Na}$ 开始上升,但在去极化未达阈电位水平时,$g_K$ 仍大于 $g_{Na}$ 而 $g_K$ 是导致膜向超极化方向变化的因素,故膜电位终将恢复至静息膜电位水平,所以仅以出现局部反应而告终。当去极化达阈电位时,$g_{Na}$ 一旦增加至等于和大于 $g_K$ 时,因 $g_{Na}$ 是导致去极化的因素,而去极化的增大,又会进一步导致 $g_{Na}$ 的上升,$g_{Na}$ 上升更加促进去极化,如此自我再生地发展,直至到达 $Na^+$ 的平衡电位时为止。这一过程即是 Na 电流的活化。当它达到顶峰时,即使膜电位仍被钳在该水平不变,也立即迅速变小,直至恢复到静息水平。此过程即是 Na 电流的灭活。从图 2.17 可以看出,从局部电位至动作电位虽似经历了突变,但膜 Na 与 K 电导的变化却是连续的,只是以 $g_{Na}=g_K$ 为界,一侧为被动的局部电位,另侧为自我再生的动作电位。

局部电位是细胞受到阈下刺激时,细胞膜两侧产生的微弱电变化(较小的膜去极化或超极化反应)。或者说是细胞受刺激后去极化未达到阈电位的电位变化。阈下刺激使膜通道部分开放,产生少量去极化或超极化,故局部电位可以是去极化电位,也可以是超极化电位。局部电位在不同细胞上由不同离子流动形成,而且离子是顺着浓度差流动,不消耗能量。局部电位具有下列特点。

(1) 等级性。指局部电位的幅度与刺激强度正相关,而与膜两侧离子浓度差无关,因为离子通道仅部分开放无法达到该离子的电平衡电位,因而不是"全或无"式的。

(2) 可以总和。局部电位没有不应期,一次阈下刺激引起一个局部反应虽然不能引发动作电位,但多个阈下刺激引起的多个局部反应如果在时间上(多个刺激在同一部位连续给予)或空间上(多个刺激在相邻部位同时给予)叠加起来(分别称为时间总和或空间总和),就有可能导致膜去极化到阈电位,从而爆发动作电位。

(3) 电紧张扩布。局部电位不能像动作电位向远处传播,只能以电紧张的方式,影响附

近膜的电位。电紧张扩布随扩布距离增加而衰减。

动作电位的形成如图 2.18 所示。当膜电位超过阈电位，能引起 Na 通道大量开放而放电动作电位的临界膜电位水平。有效刺激本身可以引起膜部分去极化，当去极化水平达到阈电位时，便通过再生性循环机制而正反馈地使 $Na^+$ 通道大量开放。

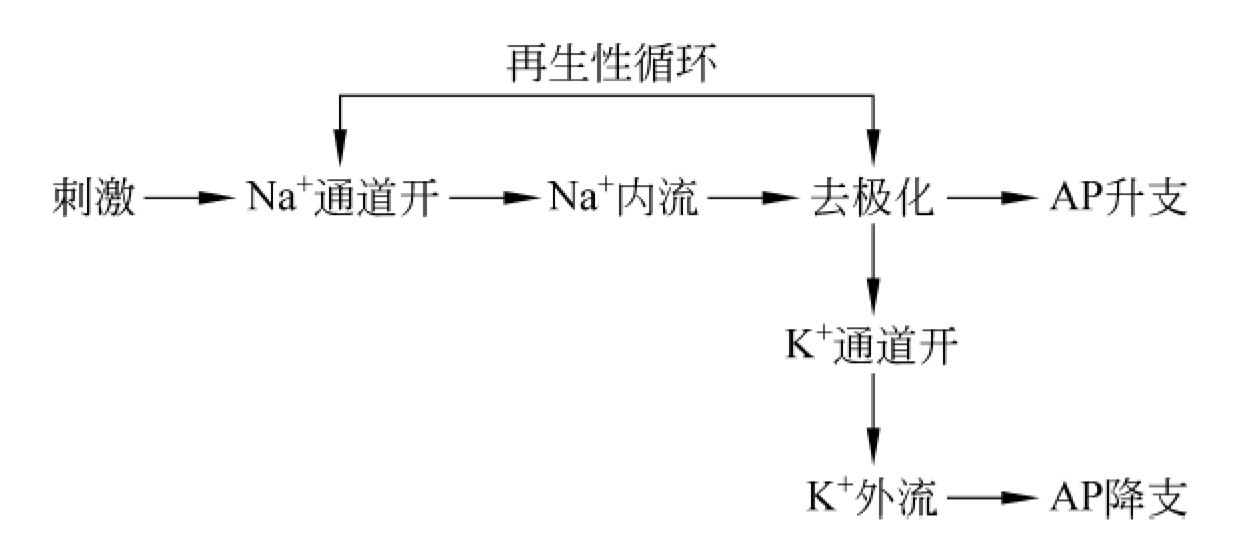



图 2.18　动作电位的形成

在膜的已兴奋区与相邻接的未兴奋区之间，由于存在电位差而产生局部电流。局部电流的强度数倍于阈强度，并且局部电流对于未兴奋区是可以引起去极化膜方向，因此，局部电流是一个有效刺激，使未兴奋区的膜去极化达到阈电位而产生动作电位，实现动作电位的传导。兴奋在同一细胞上的传导，实际上是由局部电流引起的逐步兴奋过程(如图 2.19 所示)。

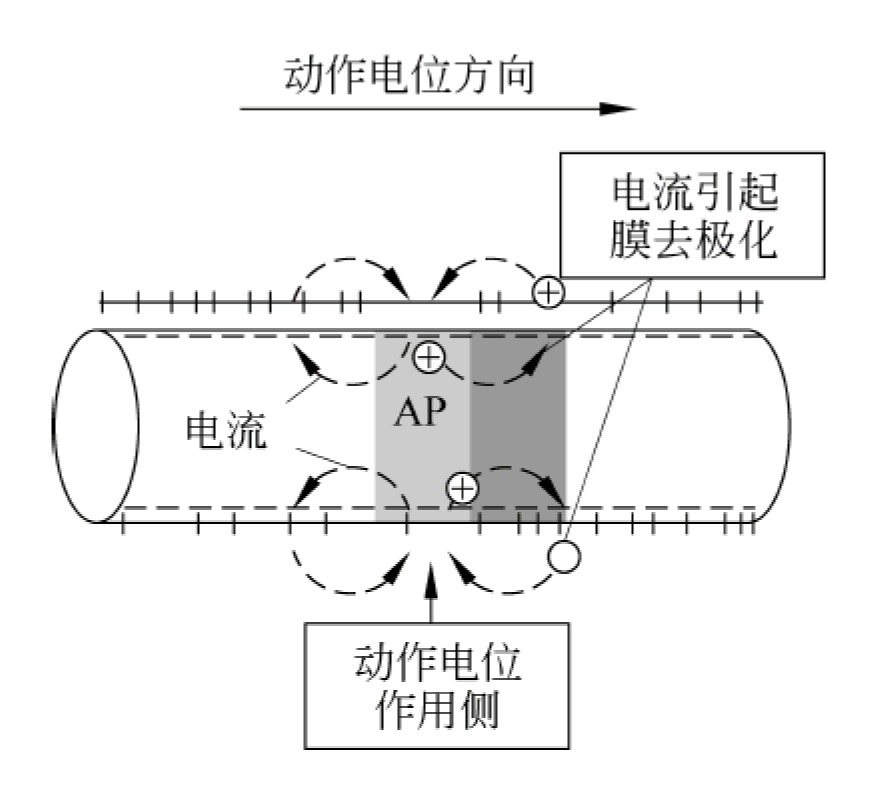



图 2.19　动作电位的传导的局部电流示意图

神经冲动是指沿神经纤维传导着的兴奋。实质是膜的去极化过程，以很快速度在神经纤维上的传播，即动作电位的传导。感受性冲动的传导，按神经纤维的不同，有两种情况：一种是无髓纤维的冲动传导，当神经纤维的某一段受到刺激而兴奋时，立即出现锋电位，即该处的膜电位暂时倒转而去极化(内正外负)，因此在兴奋部位与邻近未兴奋部位之间出现了电位差，并发生电荷移动，称为局部电流，这个局部电流刺激邻近的安静部位，使之兴奋，即产生动作电位，这个新的兴奋部位又通过局部电流再刺激其邻近的部位，依次推进，使膜的锋电位沿整个神经纤维传导；另一种是有髓神经纤维的冲动传导，其传递是跳跃性的。早在 1871 年郎飞(Ranviar)就发现外周有髓鞘纤维的髓鞘不是连续地包在轴突外面，而是有规律地每隔 1～2mm 便中断一次。后人便将髓鞘中断处称为郎飞结。关于它的生理功能长期未被阐明，1925 年利利(Lillie)曾根据在以金属丝模拟神经纤维传导的实验基础上提出设想，认为神经兴奋可能是从郎飞结到郎飞结进行跳跃式传导。有髓神经纤维的髓鞘有电绝缘性，局部电流只能产生在两个郎飞结之间，称为跳跃传导(如图 2.20 所示)。

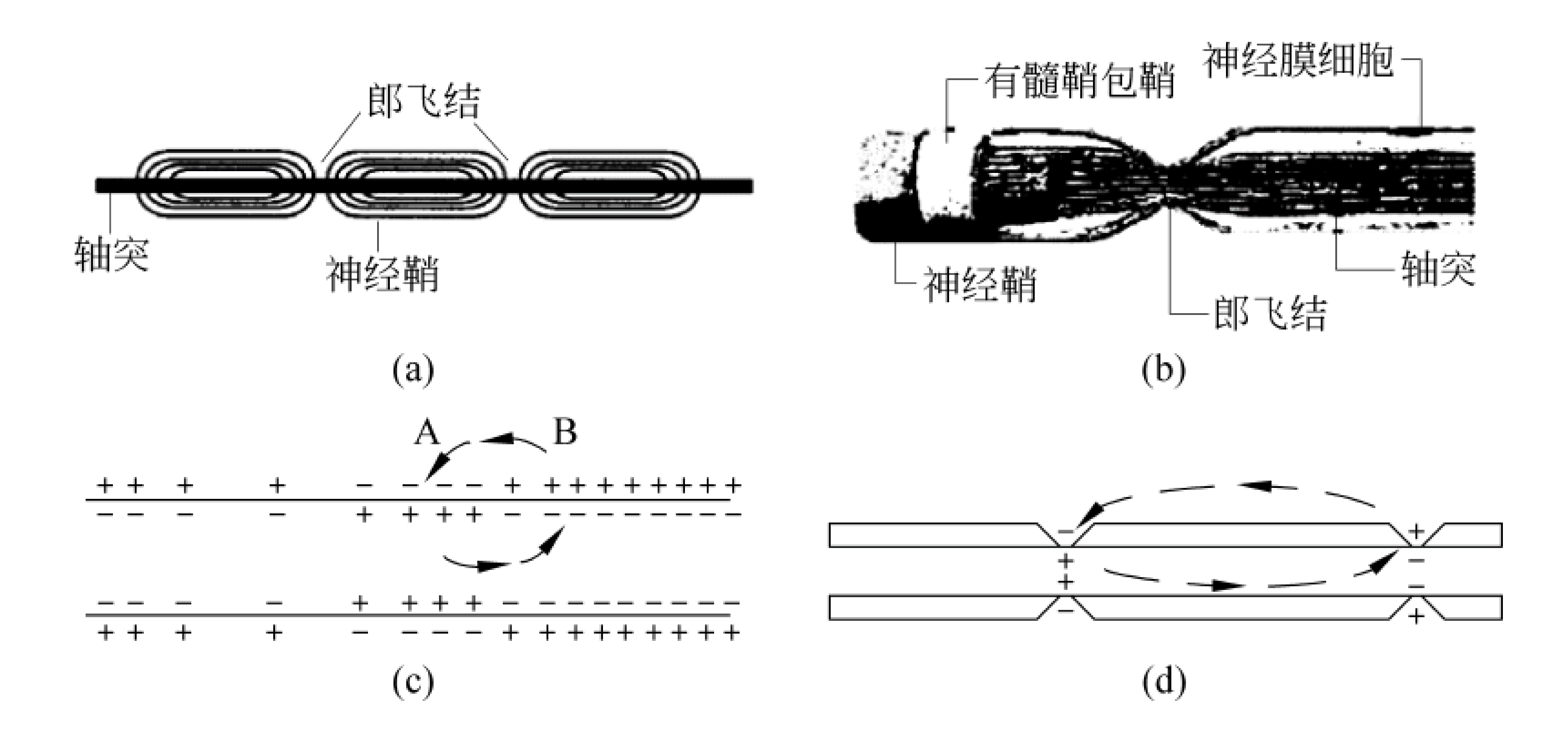



图 2.20 动作电位跳跃传导示意图

有髓神经纤维因有髓鞘,使离子不能有效地通过,但在郎飞结处轴突裸露,此处膜的通透性比无髓纤维膜的通透性大 500 倍左右,离子很容易通透,因而当一郎飞结处兴奋时,这一区域出现除极,局部电流只能沿轴突内部流动,直至下一个未兴奋的郎飞结处才穿出。在局部电流的刺激下,兴奋就以跳跃方式从一个郎飞结传至下一个郎飞结而不断向前传导。所以,有髓纤维的传导速度比无髓纤维更快。神经冲动的传递有以下特征:完整性,即神经纤维必须保持解剖学上与生理学上的完整性;绝缘性,即神经冲动在传导时不能传导至同一个神经干内的邻近神经纤维;双向传导,即刺激神经纤维的任何一点,产生的冲动可沿纤维向两端同时传导;相对不疲劳性和非递减性。

## 2.8 离子通道

1991 年的诺贝尔生理学奖授予给了尼赫(E Neher)和萨克曼(B Sakman),因为他们的重大成就——细胞膜上单离子通道的发现。细胞是通过细胞膜与外界隔离的,在细胞膜上有很多通道,细胞就是通过这些通道与外界进行物质交换的。这些通道由单个分子或多个分子组成,允许一些离子通过。通道的调节影响到细胞的生命和功能。1976 年尼赫和萨克曼合作,用新建立的膜片钳技术成功地记录了 nAc 恤单离子通道电流,开创了直接实验研究离子通道功能的先河。结果发现当离子通过细胞膜上的离子通道的时候,产生十分微弱的电流。尼赫和萨克曼在实验中,利用与离子通道直径近似的钠离子或氯离子,最后达到共识:离子通道是存在的,以及它们是如何发挥功能的。有一些离子通道上有感应器,他们甚至发现了这些感受器在通道分子中的定位,如图 2.21 所示。

1981 年英国的米勒迪(R Miledi)研究室将生物合成 nAchR 的 cRNA 注射到处于一定发育阶段(阶段Ⅴ)的非洲爪瞻卵细胞中,成功地在其膜中表达了该离子通道型受体。在 1983—1984 年日本的 Numa 研究室又利用重组 DNA 克隆技术首次确定了分子量达 20 余万的电鱼器官的 nAchR 和 Na 通道的全一级结构。上述三项工作不仅从功能和结构上直接证明了离子通道的存在,也为分析离子通道的功能与结构提供了有效的研究方法。

在神经元膜中已发现了 12 种以上基本类型的离子通道,每种类型又有一些相近的异构体。离子通道可在多种构象之间转换,但从是否容许离子通过其微孔道的现象看,只有开放

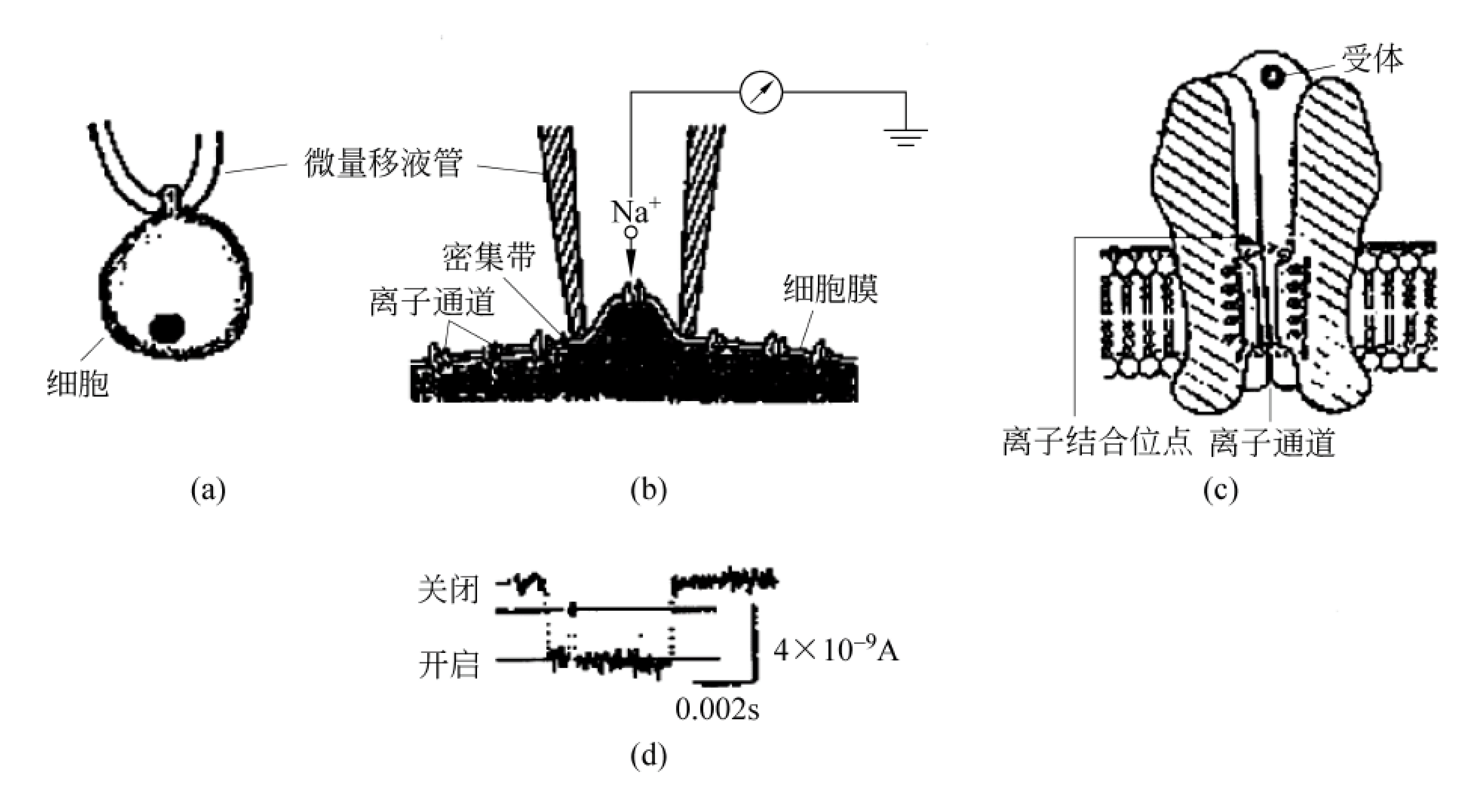



图 2.21　离子通道示意图

和关闭两种状态。离子通道在开放和关闭之间的转换是由其微孔道的闸门控制的，这一机制被称为闸控[254]。实际上在多种离子通道，如 Na 通道除开放和关闭之外，起码尚有一个被称为灭活的关闭态。

关于闸控的机制尚不十分清楚，曾设想三种方式：①孔道内的一处被闸住（如电压门控 Na 通道和 K 通道）；②全孔道发生结构变化封住孔道（如缝隙连接通道）；③由特殊的抑制粒子将通道口塞住（如电压门控 K 通道）[358]。已知有电压、机械牵拉和化学配基这三类动因可调控通道闸门的活动，相应的离子通道便被分别称为电压门控、机械门控和配基门控离子通道。

离子通道是胞膜的结构蛋白中的一类。它们贯穿胞膜并分散地存在于膜中。自从 Numa 研究室首次以 DNA 克隆技术确定了电鱼电器官的 nAchR 和 Na 通道的全氨基酸序列以来，已阐明了多种离子通道的一级结构，再加上由 X 光衍射、电子衍射和电镜技术等所得资料，已有可能对它们的二级结构、分子中的功能基团及其进化与遗传等进行判断与分析。

根据已有关于离子通道一级结构的资料，可将编码它们的基因分为三个家族，因为每个家族成员都有极为相似的氨基酸序列，从而它们被认为是由共同先祖基因演化而来：①编码电压门控 Na、K 和 Ca 通道基因家族；②编码配基门控离子通道基因家族，此族成员中有由 Ach、GABA、甘氨酸或谷氨酸激活的离子通道；③编码缝隙连接通道的基因家族。

# 2.9　脑电信号

## 2.9.1　脑电信号分类

脑电分为自发脑电（Spontaneous Electroencephalogram，EEG）和诱发脑电（Evoked Potential，EP）两种。自发脑电是指在没有特定的外加刺激时，人脑神经细胞自发产生的电位变化。这里，所谓“自发”，是相对的，指的是没有特定外部刺激时的脑电。自发脑电是非

平稳性比较突出的随机信号,不但它的节律随着精神状态的变化而不断变化,而且在基本节律的背景下还会不时地发生一些瞬念,如快速眼动等。诱发脑电是指人为地对感觉器官施加刺激(光的、声的或电的)所引起的脑电位的变化。诱发脑电按刺激模式可分为听觉诱发电位(Auditory Evoked Potential,AEP)、视觉诱发电位(Visual Evoked Potential,VEP)、体感诱发电位(Somatosensory Evoked Potential,SEP),以及利用各种不同的心理因素如期待、预备,以及各种随意活动进行诱发的事件相关电位(Event Related Potentials,ERP)等。事件相关电位把大脑皮质的神经生理学与认知过程的心理学融合了起来,它包括 P300(反映人脑认知功能的客观指标)、N400(语言理解和表达的相关电位)等内源性成分。ERP 和许多认知过程,如心理判断、理解、辨识、注意、选择、做出决定、定向反应和某些语言功能等有密切相关的联系。

自发脑电信号反映了人脑组织的电活动及大脑的功能状态,其基本特征包括周期、振幅、相位等。关于 EEG 的分类,国际上一般按频带、振幅不同将 EEG 分为下面几种波。

(1) δ波:频带范围 0.5~3Hz,振幅一般在 100μV 左右。在清醒的正常人的脑电图中,一般记录不到δ波。在成人昏睡时,或者存婴幼儿和智力发育不成熟的成人上,可以记录到这种波。在受某些药物影响时,或大脑有器质性病变时也会引起δ波。

(2) θ 波:频带范围 4~7Hz,振幅一般为 20~40μV,在额叶、顶叶较明显,一般困倦时出现,是中枢神经系统抑制状态的表现。

(3) α 波:频带范围 8~13Hz,节律的波幅一般为 10~40μV,正常人的 α 波的振幅与空间分布,也存在着个体差异。α 波的活动在大脑各区都有,不过以顶枕部最为显著,并且左右对称,安静及闭眼时出现最多,波幅亦最高,睁眼、思考问题时或接受其他刺激时,α 波消失而出现其他快波。

(4) β 波:频带范围 14~30Hz,振幅一般不超过 30μV,分布于额、中央区及前中颞,在额叶最容易出现。生理反应时 α 节律消失,出现 β 节律。β 节律与精神紧张和情绪激动有关。所以,通常认为 β 节律属于“活动”类型或去同步类型的。

(5) γ 波:频带范围 30~45Hz,振幅一般不超过 30μV,额区及中央最多,它与 β 同属快波,快波增多,波幅增高是神经细胞兴奋型增高的表现。

通常认为,皮质病变会引起一些脑波中异常频率成分,正常人的脑波频率范围一般在 4~45Hz 之间。

事件相关电位把大脑皮质的神经生理学与认知过程的心理学融合了起来,和许多认知过程,如心理判断、理解、辨识、注意、选择、做出决定、定向反应和某些语言功能等有密切相关的联系。典型的事件相关电位如下。

(1) P300:P300 是一种事件相关电位,其峰值大约出现在事件发生后 300ms,相关事件发生的概率越小,所引起的 P300 越显著。

(2) 视觉诱发电位(VEP):视觉器官受到光或图形刺激后,在大脑特定部位所记录的 EEG 电位变化,称之为视觉诱发电位。

(3) 事件相关同步(ERS)或去同步电位(ERD):单边的肢体运动或想像运动,对侧脑区产生事件相关去同步电位,同侧脑区产生事件相关同步电位。

(4) 皮层慢电位:皮层慢电位(SCP)是皮层电位的变化,持续时间为几百毫秒到几秒,实验者通过反馈训练学习,可以自主控制 SCP 幅度产生正向或负向偏移。

采用以上几种脑电信号作为 BCI 输入信号，具有各自的特点和局限。P300 和 VEP 都属于诱发电位，不需要进行训练，其信号检测和处理方法较简单且正确率较高，不足之处是需要额外的刺激装置提供刺激，并且依赖于人的某种知觉（如视觉）。其他几类信号的优点是不依赖外部刺激就可产生，但需要大量的特殊训练。

## 2.9.2 脑电信号分析

1932 年，第耶(G Dietch)首先用傅里叶变换进行了 EEG 特征分析之后，在 EEG 研究领域中相继引入了频域分析、时域分析等经典方法。近年来，小波分析、非线性动力学分析、神经网络分析、混沌分析、统计学等方法以及各种分析方法的有机结合，有力地推动了 EEG 信号分析方法的发展。随着研究工作的深入，结合时间和空间信息的脑电模式分析也成为脑电信号研究的一种有效的途径。目前广泛应用的 EEG 信号分析技术有时域分析、频域分析、时频分析和时空分析。

(1) 时域分析：直接从时域提取特征是最早发展起来的方法，因为它直观性强，物理意义比较明确。时域分析主要用来直接提取波形特征，如过零截点分析、直方图分析、方差分析、相关分析、峰值检测及波形参数分析、相干平均、波形识别等。另外，利用参数模型（如 AR 模型等）提取特征，也是信号时域分析的一种重要手段，这些特征参数可用丁 EEG 的分类、识别和跟踪等。然而，由于脑电信号的波形过于复杂，目前还没有一个特别行之有效的 EEG 波形分析方法。

(2) 频域分析：由于 EEG 信号的很多主要特征是反映在频域上的，功率谱估计是频域分析的重要手段，因此谱分析技术在脑电信号处理中占有特别重要的位置。它的意义在于把幅度随时间变化的脑电波变换为脑电功率随频率变化的谱图，从而可直观地观察到脑电频率的分布与变换情况。谱估计法一般可分为经典方法与现代方法，经典的谱估计方法是直接按定义用有限长数据来估计，即以短时间段数据的傅里叶变换为基础的周期法，主要有两种途径：间接法，先估计相关函数，再经过傅里叶变换得到功率谱估计（根据维纳-辛钦定理）；直接法，对随机数据直接进行傅里叶变换，然后取其幅值的平方得到相应的功率谱估计，又称为周期图法。这两种方法存在的共同问题是估计的方差特性不好，而且估计值沿频率轴的起伏比较剧烈，数据越长，这种现象越严重。为改善谱估计的分辨率，以参数模型为基础形成了一套现代谱估计理论。参数模型估计方法对数据处理能得到高分辨率的谱分析结果，为 EEG 信号频域特征的提取提供了有效手段。但是，功率谱估计不能反映出脑电频谱的时变性。所以，对脑电这样的时变非平稳过程单从频域的功率谱估计会丢失时变的信息。

(3) 时频分析：信号的时频分析技术，不同于以往的单纯时域或者频域分析，它是一种同时在时间和频率域中对信号进行分析的技术，主要分为线性变化和非线性变换两类。线性变换主要包括：短时傅里叶变换、Gabor 变换和小波变换技术。非线性变换主要包括 Wigner-Ville 分布、Cohen 类分布等。时频分析的主要思想是把时域信号在时间-频率平面中展开，将以时间为自变量的信号表示成以时间和频率两个参数为自变量的函数，从而表现出信号不同时间点的频率成分。与传统的傅里叶分析相比，时频分析更加有利于表现非平稳信号和时变信号的特征，突出信号的瞬态特征。在脑电信号分析中，主要应用时频分析技术进行 EEG 特征波形识别和特征提取。目前应用最为广泛的方法是小波变换理论。小波

分析在高频时使用短窗口,而在低频时使用宽窗口,充分体现了相对带宽频率分析和适应变分辨率分析的思想,从而为信号的实时分析提供了一条可能途径。目前,脑电信号的时频分析研究已取得了很多有价值的研究成果。

(4) 时空分析:考虑脑电在头皮的空间分布,将时间和空间的信息进行融合分析的时空分析方法有利于揭示和增强多导脑电信号中的隐含特征。例如,运动、感知、认知等活动在空间上的表现部位有明显的差别,因此,将时间和空间的信息进行融合分析并识别就有可能得到更加深入的研究结果。时空模式的分析方法比较多,如微状态、空间谱估计、经典统计方法(相关函数、互相关函数)、空间滤波器等。其中结合多维统计分析方法的空间滤波方法,如主成分分析(Principal Component Analysis,PCA)、独立分量分析(Independent Component Analysis,ICA)、公共空间模式(Common Spatial Pattern,CSP),在脑电信号分析处理领域都得到了非常重要的应用。具体来说,PCA 是一种线性变换,处理过程就是对信号做奇异值分解,然后找出信号中的主要成分来作为判断的依据;而基于高阶统计量的 ICA,代表着现代统计信号分析理论的最新发展,研究表明,ICA 非常适合多导 EEG 信号的分析处理,在脑电消噪和特征提取等方面取得了很好的效果;通过计算空间滤波来检测事件相关去同步现象(ERD)的 CSP 算法,是目前最成功的进行脑电信号特征提取算法之一,已被广泛应用在 BCI 中。时空分析方法能给人们提供更多的信息,是近年来 EEG 信号分析中的一个重要研究方向。

## 2.10 神经系统

神经系统是机体各种活动的"管理机构"。它通过分布在身体各部分的许多感受器和感觉神经获得关于内、外环境变化的信息;经过各级中枢的分析综合,发出信号来控制各种躯体结构和内脏器官的活动。

神经系统按其形态和所在部位可分为中枢神经系统和周围神经系统。中枢神经系统包括位于颅腔内的脑和位于椎管内的脊髓。神经系统按其性质又可分为躯体神经和内脏神经。躯体神经的中枢部分在脑和脊髓内,周围部分参与构成脑神经和脊神经。躯体感觉神经通过其末梢的感受器,接受来自皮肤、肌肉、关节、骨等处的刺激,并将冲动传入中枢;躯体运动神经传导发自中枢的运动冲动,通过效应器使骨骼肌随意收缩与舒张。内脏神经的中枢部分也在脑和脊髓内,其周围部分除随脑神经和脊神经走行外,还有较独立的内脏神经周围部分。内脏运动神经又分为交感神经和副交感神经,管理心血管和内脏器官中的心肌、平滑肌和腺体。

### 2.10.1 中枢神经系统

中枢神经系统包括脑和脊髓两大部分。在整个中枢神经系统中,脑是最主要的部分。对个体行为而言,几乎所有的复杂活动,如学习、思维、知觉等都与脑神经有密切的关系。脑的主要构造分为后脑、中脑及前脑三大部分。每一部分又各自包括数种神经组织。

脊髓在脊柱之内,上接脑部,外联周围神经,由 31 对神经分配在两侧所构成。脊髓的主要功能如下。

(1) 负责将始自感受器传入神经送来的神经冲动,传递给脑部的高级中枢,并将脑部传来的神经冲动经由传出神经而终止于运动器官。所以,脊髓是周围神经和脑神经中枢之间的通路。

(2) 接受会入神经传来之冲动后,直接发生反射活动,成为反射中枢。

图 2.22 给出脊髓的横切面图。灰质呈蝴蝶形,中部有中央管,上通第四脑室,两侧灰质向前后延伸形成前角和后角,在胸腰段和骶段前后角之间还有侧角。前角联系前根,与运动有关。后角联系后根,与感觉有关。侧角是植物性神经的节前神经元。

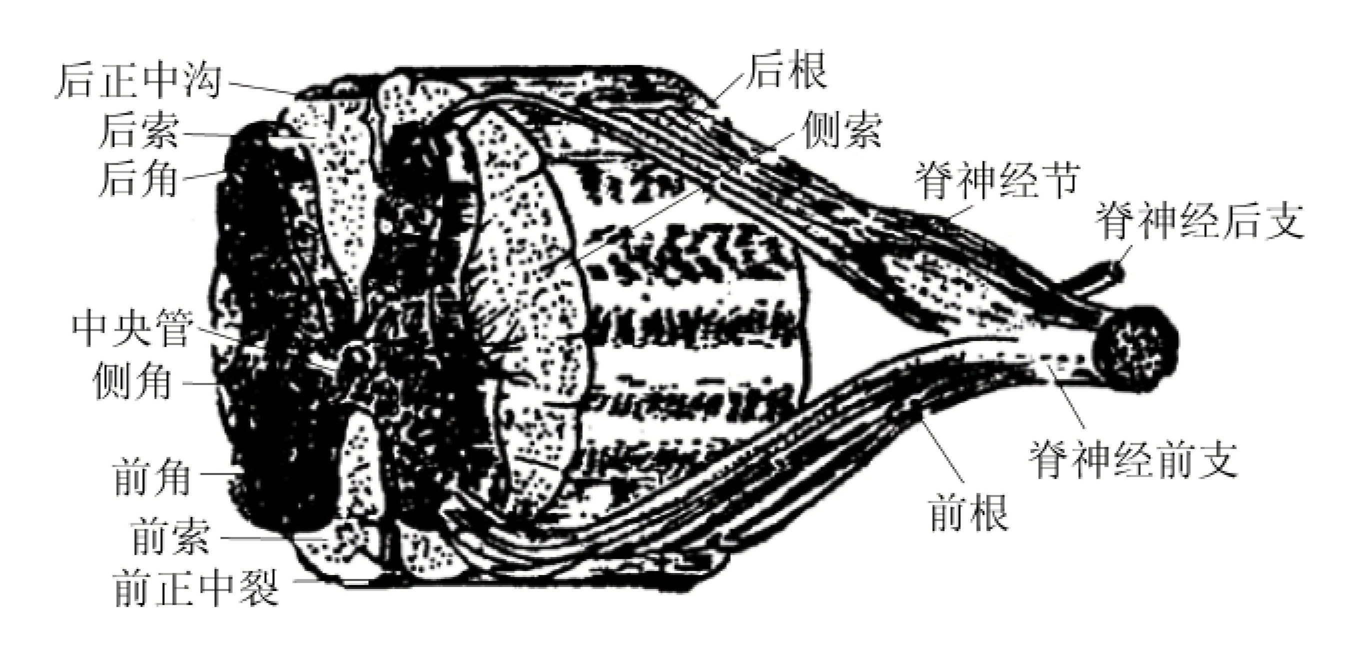



图 2.22　脊髓的横切面图解

白质在灰质周围,以前后根为界,分为前索、侧索和后索。各索内有许多上下行纤维束,是联系脑和脊髓的传导通路。紧靠灰质周围有固有束,它们是脊髓各节间的联系。

## 2.10.2　周围神经系统

周围神经系统包括体干神经系统和自主神经系统。体干神经系统,遍布于头、面、躯干及四肢的肌肉内,这些肌肉均为横纹肌。横纹肌之运动,由体干神经所支配。体干神经依其功能又分为传入神经和传出神经两类。传入神经与感觉器官的感受器相连接,负责把外界刺激所引起的神经冲动传递到中枢神经,所以这类神经也称为感觉神经,构成感觉神经的基本单位,即为感觉神经元。

中枢神经接受外来的神经冲动后,即产生反应。反应也是以神经冲动的形式,由传出神经将之传到运动器官,并引起肌肉的运动,所以这类神经也称为运动神经。运动神经的基本单位为运动神经元。运动神经元将中枢传出的冲动传到运动器官,产生相应的动作,作出一定的反应。

上述体干神经系统是管理横纹肌的行为。而对内部平滑肌、心肌及腺体的管理,即对内脏机能的管理是自主神经系统。内脏器官的活动与驱干肌肉系统的活动不同,有一定的自动性。人不能由意志直接指挥其内脏器官的活动,内脏的传入冲动与皮肤的或其他特殊感觉器官的传入冲动不同,往往不能在意识上发生清晰的感觉。在自主神经系统内,按其起源部位及生理功能的不同,又分为交感神经系统和副交感神经系统。

交感神经系统起源于胸脊髓和腰脊髓,接受脊髓、延髓及中脑各中枢所发出之冲动,受中枢神经系统所管制。故严格而论,不能称为自主,只是不受个体意志支配而已。交感神经主要分布于心、肺、肝、肾、脾、胃肠等内脏与生殖器官以及肾上腺等处。另一部分分布于头部及颈部之血管、体壁及竖毛肌、眼之虹膜等处。交感神经系统的主要功能为兴奋各内脏器官、腺体以及其他有关器官等。例如,当其兴奋时能使心跳加速、血压升高,呼吸量增大,血

液内糖分增加。瞳孔放大以及促进肾上腺素的分泌等;唯对唾液的分泌则有抑制作用。

副交感神经系统由部分脑神经(Ⅲ—动眼神经、Ⅶ—面神经、Ⅸ—吞咽神经、Ⅹ—迷走神经)和起源于脊髓骶部的盆神经所组成。副交感神经节接近效应器或者就在效应器内,所以节后纤维极短,通常只能看到节前纤维。副交感神经的主要功能与交感神经相反,因而对交感神经产生一种对抗作用。例如在心脏中副交感神经具有抑制作用,而交感神经具有增强其活动的作用;又如在小肠中,副交感神经具有增强其运动的作用,而交感神经却具有抑制作用,其作用恰与心脏中的相反。表 2.1 给出了自主神经系统的机能。

**表 2.1　自主神经系统的机能**

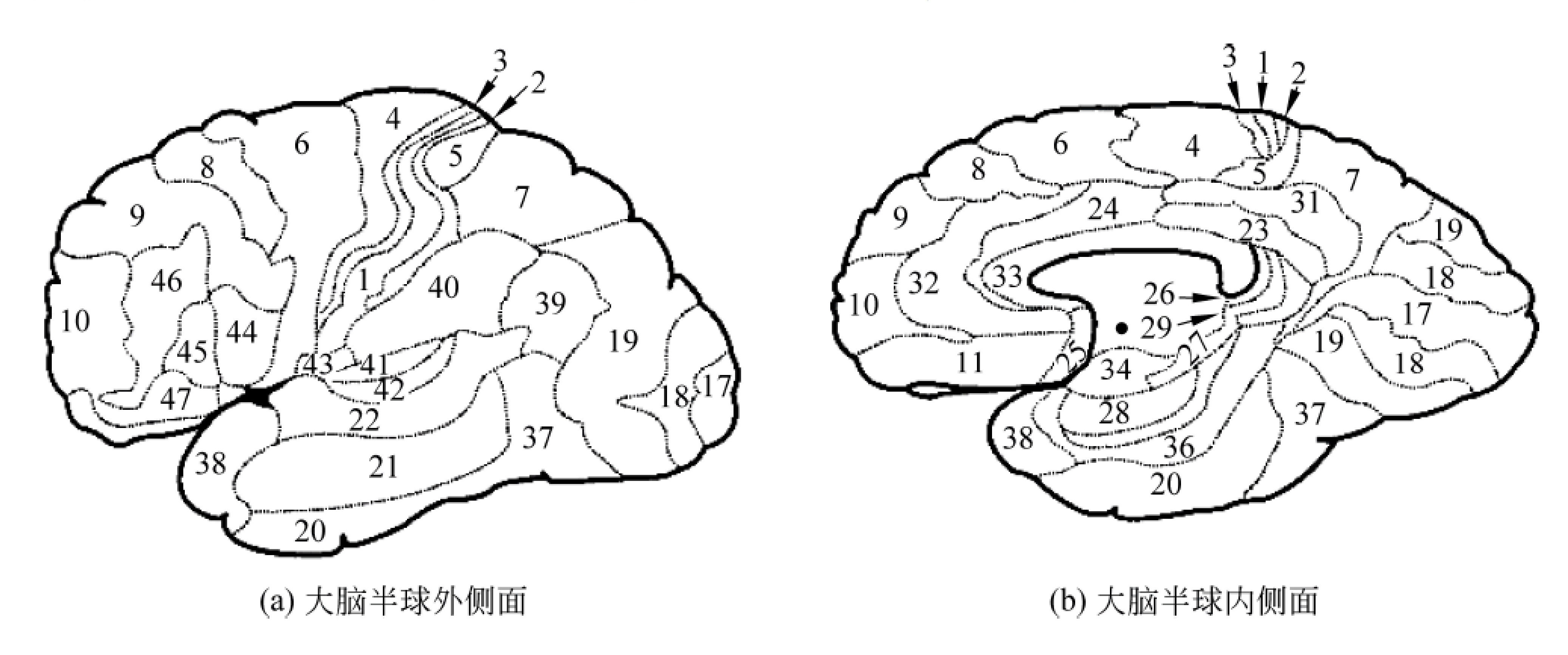
| 器官 | 交感神经系统 | 副交感神经系统 |
|---|---|---|
| 循环器官 | 心跳加快,冠状血管舒张,腹腔内脏与皮肤末梢血管收缩,储血库(如脾)收缩 | 抑制心跳,冠状血管收缩部分器官末梢血管舒张 |
| 呼吸器官 | 支气管舒张 | 支气管收缩,促进黏膜腺的活动 |
| 消化器官 | 分泌黏稠的唾液,抑制胃、肠运动,降低肠管平滑肌的紧张。抑制胆囊收缩 | 分泌稀薄唾液,肠胃肌肉的蠕动增强,紧张性升高。促进胆囊收缩 |
| 泌尿生殖器官 | 肾脏血管收缩,膀胱逼尿肌舒张,外生殖器官血管收缩,促进子宫收缩(有孕子宫)或舒张(无孕子宫) | 膀胱逼尿肌收缩,外生殖器官舒张,阴茎勃起 |
| 眼 | 瞳孔散大<br>眼宽及眼睑的平滑肌收缩(眼球外突),睫状肌舒张 | 瞳孔缩小<br>睫状肌收缩促进泪腺分泌 |
| 皮肤 | 毛囊肌收缩,汗液分泌 | |
| 代谢 | 促进异化作用,促进肾上腺素分泌 | 促进同化作用,促进腺岛素分泌 |

## 2.11　大脑皮质

1860 年法国外科医生布洛卡(P Broca)观察了一个病例,这位病人可以理解语言,但不能说话。他的喉、舌、唇、声带等都没有常规的运动障碍。他可以发出个别的词和哼曲调,但不能说完整的句子,也不能通过书写表达他的思想。尸体解剖发现,病人大脑左半球额叶后部有一鸡蛋大的损伤区,脑组织退化并与脑膜粘连,但右半球正常。布洛卡后来研究了 8 个相同的病人,都是在大脑左半球这个区域受损。这些发现使布洛卡在 1864 年宣布了一条著名的脑机能的原理:“我们用左半球说话”。这是第一次在人的大脑皮质上得到机能定位的直接证据。现在把这个区(Brodmann44、45 区)叫做布洛卡表达性失语症区,或布洛卡区。这个控制语言的运动区只存在于大脑左半球皮层,这也是人类大脑左半球皮层优势的第一个证据。

1870 年两位德国生理学家弗里奇(G Fritsch)和希齐格(E Hitzig)发现,用电流刺激狗大脑皮质的一定部位,可以规律性地引起对侧肢体一定的运动。这是第一次用实验证明了大脑皮质上存在不同的机能定位。后来韦尔尼克(C Wernicke)又发现另一个与语言能力有关的皮层区,现在叫做韦尔尼克区,是在颞叶的后部与顶叶和枕叶相连接处。这个区受损伤的病人可以说话但不能理解语言,即可以听到声音,却不能理解它的意义。这个区也是在左半球得到更加充分的发展。

从 19 世纪以来经过生理学家、医生等多方面的实验研究和临床观察，以及把临床观察、手术治疗和科学实验结合进行，得到了关于大脑皮质机能的许多知识。20 世纪 30 年代彭菲尔德等对人的大脑皮质机能定位进行了大量的研究。他们在进行神经外科手术时，在局部麻醉的条件下用电流刺激病人的大脑皮质，观察病人的运动反应，询问病人的主观感觉。布洛德曼根据细胞构筑的不同，将人的大脑皮质分成 52 区（图 2.23）。从功能上来分，大脑皮质由感觉皮层、运动皮层和联合皮层组成。感觉皮层包括视皮层（17 区）、听皮层（41、42 区）、躯体感觉皮层（1、2、3 区）、味觉皮层（43 区）和嗅觉皮层（28 区）；运动皮层包括初级运动区（4 区）、运动前区和辅助运动区（6 区）；联合皮层包括顶叶联合皮层、颞叶联合皮层和前额叶联合皮层。联合皮层不参与纯感觉或运动功能，而是接受来自感觉皮层的信息并对其进行整合，然后将信息传至运动皮层，从而对行为活动进行调控。联合皮层之所以被这样称呼，就是因为它在感觉输入与运动输出之间起着联合的作用。



(a) 大脑半球外侧面

(b) 大脑半球内侧面

图 2.23　人类大脑皮质分区

人类顶叶联合皮层包括 Brodmann 5、7、39 和 40 区。5 区主要接受初级躯体感觉皮层（1、2、3 区）和丘脑后外侧核的投射，而 7 区主要接受纹状前视区、丘脑后结节、颞上回、前额叶皮层和扣带回（23、24 区）的投射。5 区和 7 区尽管有着不同的输入来源，但却有着共同的投射靶区，这些靶区包括运动前区、前额叶皮层、颞叶皮层、扣带回、岛回和基底神经节。不同的是，5 区更多地投射到运动前区和运动区，而 7 区投射到那些与边缘结构有联系的颞叶亚区（5 区则没有这种投射）。此外，7 区还直接向旁海马回投射，并接受来自蓝斑和缝际核的投射。因此，5 区可能更多地参与躯体感觉信息及运动信息的处理，7 区则可能主要参与视觉信息处理，并参与运动、注意和情绪调节等功能。

人类前额叶联合皮层由 Brodmann 9～14 区及 45～47 区组成。11～14 区及 47 区总称为前额叶眶回；9、10、45 和 46 区总称为前额叶背外侧部，有些作者把 8 区和 4 区也归纳到前额叶皮层的范畴。前额叶联合皮层在解剖学上具有几个显著的特征：位于大脑新皮层的最前方；具有显著发达的颗粒第Ⅳ层；接受丘脑背内侧核的直接投射；具有广泛的传入传出纤维联系。动物从低等向高等进化，前额叶联合皮层面积也相应地变得越来越大。灵长类（包括人类）具有最发达的前额叶联合皮层。在人类，前额叶联合皮层占整个大脑皮质面积的 29%左右。

前额叶联合皮层有着极丰富的皮层及皮层下纤维联系。前额叶皮层与纹状前视区、颞

叶联合皮层、顶叶联合皮层有着交互的纤维联系。前额叶皮层是唯一与丘脑背内侧核有交互纤维联系的新皮层,也是唯一向下丘脑有直接投射的新皮层。前额叶皮层与基底前脑、扣带回及海马回有直接或间接纤维联系。前额叶皮层发出纤维投射到基底神经节(尾核和壳核)等等。这种复杂的纤维联系决定了前额叶皮层功能上的复杂性。

人类大脑皮质是一个极其复杂的控制系统,大脑半球表面的一层灰质,平均厚度 2～3mm。皮层表面有许多凹陷的"沟"和隆起的"回"。成人大脑皮质的总面积,可达 $2200cm^2$,具有数量极大的神经元,估计约为 140 亿个。其类型也很多,主要是锥体细胞、星状细胞及梭形细胞。神经元之间具有复杂的联系。但是,各种各样的神经元在皮层中的分布不是杂乱的,而是具有严格层次的。大脑半球内侧面的古皮层比较简单,一般只有如下三层。

(1) 分子层。

(2) 锥体细胞层。

(3) 多形细胞层。

大脑半球外侧面等处的新皮层,具有如下六层。

(1) 分子层,细胞很少,但有许多与表面平行的神经纤维。

(2) 外颗粒层,主要由许多小的锥体细胞和星状细胞组成。

(3) 锥体细胞层,主要为中型和小型的锥体细胞。

(4) 内颗粒层,由星状细胞密集而成。

(5) 节细胞层,主要含中型及大型锥体细胞,在中央前回的锥体细胞特别大,它们的树突顶端伸到第一层,粗长的轴突下行达脑干及脊髓,组成锥体束的主要成分。

(6) 多形细胞层,主要是梭形细胞,它们的轴突除一部分与第 5 层细胞的轴突组成传出神经纤维下达脑干及脊髓外,一部分走到半球的同侧或对侧,构成联系皮质各区的联合纤维。

从机能上看,大脑皮质 1、2、3、4 层主要接受神经冲动和联络有关神经,特别是从丘脑来的特定感觉纤维,直接进入第 4 层。第 5、6 层的锥体细胞和梭形细胞的轴突组成传出纤维,下行到脑干与脊髓,并通过脑神经或脊神经将冲动传到身体有关部位,调节各器官、系统的活动。这样大脑皮质的结构不但具有反射通路的性质,而且是各种神经元之间的复杂的连锁系统。由于联系的复杂性和广泛性,使皮层具有分析和综合的能力,从而构成了人类思维活动的物质基础。

对大脑体表感觉区皮层结构和功能的研究指出,皮层细胞的纵向柱状排列构成大脑皮层的最基本功能单位,称为功能柱。这种柱状结构的直径为 200～500μm,垂直走向脑表面,贯穿整个六层。同一柱状结构内的神经元都具有同一种功能,例如都对同一感受野的同一类型感觉刺激起反应。在同一刺激后,这些神经元发生放电的潜伏期很接近,仅相差 2～4ms;说明先激活的神经元与后激活的神经元之间仅有几个神经元接替;亦说明同一柱状结构内神经元联系环路只需通过几个神经元接替就能完成。一个柱状结构是一个传入-传出信息整合处理单位,传入冲动先进入第 4 层,并由第 4 层和第 2 层细胞在柱内垂直扩布,最后由第 3、第 5、第 6 层发出传出冲动离开大脑皮质。第三层细胞的水平纤维还有抑制相邻细胞柱的作用,因此一柱发生兴奋活动时,其相邻细胞柱就受抑制,形成兴奋和抑制镶嵌模式。这种柱状结构的形态功能特点,在第二感觉区、视区、听区皮层和运动区皮层中也一样存在。

# 第 3 章<<

# 神经计算

神经计算是建立在神经元模型和学习规则基础之上的一种计算范式，由于特殊的拓扑结构和学习方式，产生了多种人工神经网路，模仿人脑信息处理的机理。人工神经网络是由大量处理单元组成的非线性、大规模、自适应的动力系统。

## 3.1 概述

神经计算研究非程序的、适应性的、大脑风格的人工神经网络信息处理的本质和能力[917]。人工神经网络是由大量处理单元组成的非线性大规模自适应动力系统，具有学习能力、记忆能力、计算能力以及智能处理功能，并在不同程度和层次上模仿人脑神经系统的信息处理、存储及检索功能。它是在现代神经科学研究成果的基础上提出的，试图通过模拟大脑神经网络处理、记忆信息的方式设计一种新的机器使之具有人脑那样的信息处理能力。同时，对这种人工神经网络的研究将进一步加深对思维及智能的认识。为了模拟大脑信息处理的机理，人工神经网络具有非线性、非局域性、非定常性、非凸性等特点。人工神经网络是算法和结构统一为一体的系统，可以看作是硬件和软件的混合体。

现代神经计算开始于麦克洛奇和皮兹的先驱工作[483]。麦克洛奇是神经学家和解剖学家。他用 20 年的时间考虑神经系统对事件的表示问题。皮兹是数学天才，于 1942 年开始神经计算的研究。1943 年麦克洛奇和皮兹在一个神经建模小组上公布了他们的论文。该小组在芝加哥大学拉谢夫斯基(N Rashevsky)领导下，五年前就很活跃了。

在他们的经典论文里，麦克洛奇和皮兹结合了神经生理学和数理逻辑的研究描述了一个神经网络的逻辑分析。他们的神经元模型假定遵循有-无模型律。如果如此简单的神经元数目足够多，适当设置连接权值并且同步操作，麦克洛奇和皮兹证明这样构成的网络原则上可以计算任何可计算函数。这是一个有重大意义的结果，有了它标志着神经网络和人工智能的诞生。

麦克洛奇和皮兹 1943 年的论文影响了冯·诺依曼，使得他在 EDVAC(Electronic Discrete Variable Automatic Computer，电子离散变元自动计算机)中使用了从麦克洛奇和皮兹的神经元导出的理想化的开关延迟要素，进而改进成为 EDVIC(Electronic Numerical Integrator and Computer，电子数值积分机和计算机)。EDVIC 是第一台通用的电子计算机，于 1943 年到 1946 年在宾夕法尼亚大学电子工程摩尔学院建成。麦克洛奇和皮兹的形式神经网络理论在冯·诺依曼 1949 年在伊利诺伊大学所作的四个报告中的第二个报告中成了显著特色。

1948 年维纳的名著《控制论》(*Cybernetics*)发表，为控制、通信和统计信号处理描述了

一些重要概念[810]。1961 年该书第二版出版发行,添加了关于学习和自组织的新材料。从第二版的第二章中可以看出,维纳在主体背景下抓住了统计机制的物理意义,但是联合统计机制和学习系统的丰硕成果却留给了霍普菲尔德。

第二个重要发展是 1949 年,赫布在他的书《行为组织学》(*The Organization of Behavir*)[311]中第一次清楚说明了突触修正的生理学习规则。特别是赫布提出大脑的连接是随着生物学会不同功能任务而连续地变化的,神经组织就是由这种变化创建起来的。赫布继承了拉莫尼(Ramony)和卡贾尔早期的假设并引入自己的学习假说:两个神经元之间的可变突触被突触两端神经元的重复激活加强了。赫布的书在心理学家中有广泛的影响,但遗憾的是对工程界却影响很少。

赫布的书是学习和自适应系统的计算模型发展的灵感源泉。1956 年罗切斯特(Rochester)、霍兰德(Holland)、哈比特(Habt)、杜伯(Duba)的论文也许是用计算机模拟测试公式化的赫布学习假说的神经理论的第一次尝试,论文报告的模拟结果表明必须加上抑制才能实际工作。同一年,乌特列(Uttley)演示了带有可修改的突触的神经网络可以学习分类简单的二值模式集。乌特列引入了所谓漏电求积和点火神经元(Leaky Integrate and Fire Neuron),后来凯恩尼罗(Caianiello)对它进行了形式分析。在更晚一些的工作中,乌特列假设了神经系统可变突触的效果依赖于突触两端波动状态的统计关系,因此和香农的理论有了联系。

在 1952 年,阿希比(Ashby)的书《大脑设计:自适应行为的起源》出版[34],这本书引人注目的是其基本观点,即自适应行为不是与生俱来而是后天习得,通过学习动物(系统)的行为变得更好。这本书强调活的生物如同机器的动态观点和有关稳定性的概念。

1954 年明斯基在普林斯顿大学写了关于神经网络的博士论文,题目是“类神经增强式系统的理论及其在大脑建模中的应用”。1961 年明斯基出版了早期关于 AI 的优秀论文“人工智能进展”。后面这篇文章包括了关于现在称为神经网络内容的很大一节。1967 年明斯基出版了《计算:有限和无限机器》这本书,第一次以书的形式扩展了麦克洛奇和皮兹的结果,并把它们放在自动理论和计算理论的背景中。

1954 年,作为通信理论先驱者之一和全息照相技术的发明者,盖博尔(Gabor)提出了非线性自适应滤波器的思想,并且他与合作者一起建立了这样的机器。随机过程产生的样本以及希望机器产生的目标函数一起提供给机器完成在线学习。

20 世纪 50 年代,泰勒(Taylor)开始研究联想记忆。接着施泰恩布什(Steinbuch)介绍了学习矩阵,这个矩阵由插在成行的“感觉”接受器和“马达”效应器之间的开关平面网络构成。在 1969 年,威尔肖(Willshaw)、布尼曼(Buneman)和洛埃特-希斬斯(Longuet-Higgins)发表了关于非全息照相的联想记忆的优秀论文。这篇文章给出了两类网络模型:实现相关矩阵的简单光学系统和由光学记忆提出的与之相关的神经网络。联想记忆早期发展的重要贡献包括安德森(Andelson)、科霍南(Kohonen)和纳卡诺(Nakano)于 1972 年的文章,他们在外积学习规则的基础上独立地引入相关矩阵记忆的思想。

冯·诺依曼是 20 世纪前 50 年的科学巨匠。为了纪念他,人们把设计数字计算机的基础命名为冯·诺依曼结构。耶鲁大学邀请他在 1956 年作 Silliman 报告。他于 1957 年去世,稍后他未完成的 Silliman 报告于 1958 年出版成书:《计算机和大脑》(*The Computer and the Brain*)。这本书很有意义,书中显示出冯·诺依曼开始意识到大脑和计算机的巨大

差异。

神经网络中一个值得特别关心的问题是利用被认为不可靠的神经元分量构建可靠的神经网络。1956 年冯·诺依曼利用冗余的思想解决了这个重要问题，基于这种思想维诺格拉德(Winograd)和考恩(Cowan)于 1963 年建议在神经网络中使用分布冗余表示，这表明了大量的元素怎样集体表示具有鲁棒性和并行性的单个概念。

在麦克洛奇和皮兹的经典论文发表 15 年以后，罗森勃拉特在他有关感知器的研究著作中提出了模式识别问题的新方法，一种新的监督学习方法[632]。感知器收敛定理使罗森勃拉特的工作取得圆满的成功。他于 1960 年提出感知器收敛定理的第一个证明。该定理的证明也出现在 Novikoff(1963)和其他人的工作中。威德罗(Widrow)和霍夫(Hoff)介绍了最小平均平方(Least Mean Square，LMS)算法，并用它构成了 Adaline(Adaptive Linear Element)。感知器和 Adaline 的区别在于训练过程。最早的可训练的具有多个自适应元素的分层神经网络之一是由威德罗和他的学生提出的 Madaline(Multiple-Adaline)结构[803]。1967 年甘利俊一(Amari)将统计梯度方法用于模式分类。1965 年 Nillson 出版《学习机》(*Learning Machine*)一书[540]，迄今为止它仍是关于用超平面区分线性可分模式的最好的作品。在 20 世纪 60 年代，感知器神经网络好像可以做任何事。明斯基和佩珀特(Papert)在 1969 年合写的《感知器》(*Perceptrons*)一书中利用数学理论证明单层感知器所能计算的本质局限[508]。在有关多层感知器的简要介绍一节中，他们认为没有任何理由假定单层感知器的任何局限可以在多层的情况下被克服。在多层感知器的设计中面临一个重要的问题就是置信度问题(即隐藏神经元在网络中的置信度问题)。明斯基于 1961 年在他的“加强学习系统的置信度问题”中首次使用了“置信度”术语。在之后的 60 年代里，解决感知器的置信度问题所必需的大多数思想和基本概念已经公式化了，也就是现在称之为霍普菲尔德网络的递归(吸引子)网络所固有的许多思想。然而，直到 20 世纪 80 年代这些基本问题求解结果才出现。

在 20 世纪 70 年代这些因素以这种或那种方式阻碍了人们进一步研究神经网络。除了一些心理学和神经学方面的专家之外，许多研究人员在那个时期都改变了研究领域，只有屈指可数的早期开创者继续神经网络研究。从工程学的角度，可将 20 世纪 70 年代追溯为神经网络的潜伏期。

在 20 世纪 70 年代一个重要的成果就是出现了利用竞争学习的自组织理论。马尔斯伯格(von der Malsburg)1973 年完成的计算机模拟工作也许是第一次演示了自组织。在人脑中拓扑序映射启发下，1976 年威尔肖和马尔斯伯格发表了第一篇关于自组织映射形成的论文。

在 20 世纪 80 年代关于神经网络的理论和设计方面取得了进展，随之神经网络的研究工作进入了恢复期。格罗斯伯格(S Grossberg)基于他的竞争学习理论的早期工作[292,293,294]，建立了一个新的自组织原则，就是著名的自适应共振理论(Adaptive Resonance Theory，ART)。基本上说，这个理论包括一个由底向上的识别层和一个由顶向下的产生层。如果输入形式和已学习的反馈形式匹配，一个叫做自适应共振的不定状态(即神经活动的放大和延长)发生了。这个前向或反向映射原已由其他的研究者在不同的条件下重新发现。

在 1982 年，霍普菲尔德基于能量函数的思想提出了一种对称连接的递归网络计算的新

方法。并且他在这种递归网络和统计物理中使用的 Ising 模型之间建立了同构。这个类推为一连串的物理理论进入神经元模型奠定了基础,因此改变了神经网络的研究方法。这类具有反馈的特殊神经网络在 80 年代引起了大量的关注,在那个时期产生了著名的 Hopfield 网络。尽管 Hopfield 网络不可能是真正的神经生物系统模型,然而它们包涵的原理,即在动态的稳定网络中存储信息的原理,是极深刻的。事实上,这个原理可以追溯到许多其他研究者的早期工作。

格雷盖(Gragg)和汤姆泼勒(Tamperley)分别于 1954 年和 1955 年观察得出正是由于神经元能被“点火”(激活)或“不点火”(静止),所以在一个网格中的原子可以用它们自旋指向“上”或“下”。

1967 年考恩引入了“sigmoid”的激活特征和一个神经元基于 Logist 函数的平滑激活条件。

格罗斯伯格于 1967—1968 年引入了一个神经元的相加模型,涉及非线性差分/微分方程,并且探测了以短期记忆为基础的模型用途。

1972 年甘利俊一独立地引入了神经元的相加模型,并用它研究随机连接的类神经元的元素的动态行为。

威尔森(Wilson)和考恩于 1972 年推导了包括兴奋和抑制模型神经元的空间局部化的群体动力学耦合非线性微分方程。

利特勒(Little)和肖于 1975 年描述了神经元激活或不激活概率模型,并用它发展了短期记忆理论。

安德森、西弗施泰(J W Siverstein)、里兹和琼斯(R S Jones)于 1977 年提出盒中脑(Brain-State-in-Box,BSB)模型,由一个非线性动力学耦合的简单联想网络组成。

因此,毫不奇怪 1982 年霍普菲尔德的论文发表后引起了很大争论[330]。但是,该文第一次使在动态的稳定网络中存储信息的原理清楚了。霍普菲尔德表明了他对从统计力学自旋玻璃体检验具有对称连接的特殊递归网络富有洞察力,对称性设计可以保证收敛到一个稳定的条件。1983 年,科恩(M A Cohen)和格罗斯伯格建立了按内容寻址记忆的一般原则连续 Hopfield 网络是一个特例。吸引子神经网络的一个与众不同的特征是一个学习的必要量——时间出现在网络的非线性动力学中。在这个背景下 Cohen-Grossberg 的定理非常重要。

1982 年另一个重大发展是科霍南关于使用一维或二维格形结构的自组织映射研究的著作[377],这在某些方面与威尔肖和马尔斯伯格稍早的工作不同。在文献中科霍南工作在分析和应用方面比威尔肖和马尔斯伯格的模型得到了更多的注意,已经成为这一领域其他创新的评估标准。

1983 年克尔潘特里克(S Kirkpatrick)、格拉特(C D Gelatt)和维奇(M P Vecchi)描述了解决组合最优化的问题的称为模拟退火的新方法[369]。模拟退火根植于统计力学,是基于梅罗波列斯(N Metropolis)等在计算机仿真中第一次使用的一个简单技术[496]。阿克列(D H Ackley),辛顿(G E Hinton)和塞杰诺斯基(T J Sejnowski)利用模拟退火的思想发展称为玻尔兹曼(Boltzmann)机的随机机器[321],它是多层神经网络的第一个成功实现。虽然玻尔兹曼机的学习算法没有反向传播算法的计算效率高,但它证明了明斯基和佩珀特的设想是不成立的,也为尼勒(R M Neal)随后的 sigmoid 信度网络发展做了铺垫工作[518]。

sigmoid 信度网络完成了两件事：(1)学习显著改善；(2)联系了神经网络和信度网络。sigmoid 信度网络学习性能的进一步提高是沙勒(L K Saul)、雅卡拉(T Jakkolla)和约丹(M I Jordan)利用一个根植于统计力学的平均场理论作出的[628]。

巴托(A G Barto)、苏顿(R S Sutton)和安德森关于强化学习的论文发表于 1983 年[60]。虽然他们不是第一次使用强化学习(例如明斯基在他 1954 年的博士论文中考虑过它)，但这篇文章引起了大量关于强化学习及其在控制中应用的兴趣。特别地，他们证明了一个强化学习系统可以在没有有益教师的情况下学习倒立摆(即车上立一个杆)平衡。学习系统仅要求当杆对竖直方向倾斜超过一定角度或车到达轨道的端点时的失败信号。1996 年贝特色卡斯(D P Bertsekas)和茨茨克利斯(J N Tsitsiklis)出版了《神经动力学程序》(*Neurodynamics*)，这本书把强化学习和 Bellmam 的动态规划相联系，把它放在一个恰当的数学基础上。

1984 年出版了布雷滕伯格(V Braitenberg)的书《工具：合成心理学的实验》(*Vehicles: Experiments in Synthetic Psychology*)。在这本书中布雷滕伯格提出了目标导向的自组织行为原则：利用公认的基本机制的一个综合体而非由顶向下的分析是最好了解一个复杂过程的方法。在科幻小说的形式下，布雷滕伯格通过描述各种具有简单内部结构的机器说明了这个重要原则。他直接或间接研究了 20 多年的动物大脑，这启发了他描述这些机器的特性和它们的行为。

1986 年鲁梅尔哈特(D E Rumelhart)、辛顿和威廉姆斯(R J Williams)报告了反向传播算法的发展。同一年，著名的鲁梅尔哈特和麦克莱伦德(J L McClelland)主编的《并行分布处理：认知微结构的探索》(*Parallel Distributed Processing: Explorations in the Microstructure of Cognition*)一书出版[635]。这本书极大地促进了反向传播算法的使用造成重大影响，它已成为最通用的多层感知器的训练算法。事实上，反向传播学习在同一时间在其他两个地方被独立提出过。在 20 世纪 80 年代中公布反向传播算法后，发现早在 1974 年 8 月哈佛大学的韦勃斯(P J Werbos)已经在他的博士学位论文中描述了[805]。韦勃斯的博士论文是有效的反向传播模式的梯度计算的第一次描述，它可以应用于包括神经网络作为其特例的一般网络模型。反向传播的基本思想可进一步追溯到勃雷森(A E Bryson)和何(Y C Ho)的书《应用最优控制》(*Applied Optimal Control*)。在该书标题为“多阶段系统”的 2.2 节中，描述了使用 Lagrange 形式的反向传播推导。但是，最终的分析得出反向传播算法的许多荣誉应给予鲁梅尔哈特、辛顿和威廉姆斯，因为他们提出了它在机器学习的应用并且演示了它怎样工作。

1988 年林斯克(R Linsker)描述了认知网络中自组织问题的新原理[439]。原理设计成在突触连接和突触动态范围限制下最大限度地保持输入活动模式的信息。其他几个视觉研究者也做出了相似的建议。但是，是林斯克使用了根植于信息理论的抽象概念构成了极大互信息(Infomax)原理的公式。林斯克的文章重新激发了应用信息理论到神经网络中的兴趣。特别是贝尔(A J Bell)和塞杰诺斯基(T J Sejnowski)的信息理论对盲信源分离问题的应用[68]，已经促使许多研究者探索用于求解大都熟知的盲解卷积的很大一类问题的信息理论模型。

同样在 1988 年，布鲁姆海德(D S Broomhead)和洛韦(D Lowe)描述了使用径向基函数(Radial Basis Function，RBF)设计多层前馈网络的过程。RBF 提供了多层感知器的另一选

择。径向基函数的基本想法至少追溯到佩希坎罗夫(O A Bashkirov)、布雷弗曼(E M Braverman)和姆希尼克(I B Muchnik)首先提出的势函数方法[62],以及艾泽曼(M A Aizerman)、布雷弗曼和罗宗纳(L I Rozonnoer)发展的势函数理论。在1973年杜达(R O Duda)和哈特(P E Hart)的经典著作《模式分类和场景分析》(*Pattern Classification and Scene Analysis*)给出了势函数方法的一个描述。尽管如此,布鲁姆海德和洛韦的文章导致了联系神经网络设计和数值分析的中重要领域以及线性自适应滤波器的大量研究的努力。1990年波杰欧(T Poggio)和吉罗希(F Girosi)利用Tikhonov的正规化理论进一步丰富了RBF网络理论[584]。

1989年米德(C A Mead)的《模拟 VLSI 和神经系统》(*Analog VLSI and Neural Systems*)一书出版。这本书提出了从神经生物学和VLSI技术吸取的不寻常的混合概念,包括米德和他的合作者写的关于硅视网膜和硅耳蜗的章节。

在20世纪90年代早期,瓦珀尼克(V N Vapnik)和他的合作者提出了计算上强有力的并被称为支持向量机(Support Vector Machines,SVM)的一种监督学习网络,用于解决模式识别、回归和密度估计问题[780,781],新方法是基于有限样本学习理论的结果。支持向量机的一个新颖的特征就是在它们的设计中以自然的方式包含了Vapnik-Chervonenkis(VC)维数。VC维数提供了神经网络从一个样本集中学习的容量的一种度量。

现在混沌是构成物理现象的关键方面已很好地建立了。许多人提出了一个问题:在神经网络研究中混沌起关键作用吗?在生物学意义下弗里曼(W J Freeman)相信这个问题的答案是肯定的[241]。根据弗里曼神经模型,活动模式不是大脑外部所加的,而是从内部构建的。混沌动力学提供了对神经元群体内和它们之间自组织模式出现的必要条件进行描述的一个基础。

霍普菲尔德1982年的文章和鲁梅尔哈特和麦克莱伦德(J L McClelland)1986年的两卷书是对20世纪80年代神经网络的复兴最有影响的作品。神经网络从麦克洛奇和皮兹的早期岁月已走过了很长一段路,已确立了它作为根植于神经科学、心理学、数学、物理学和工程的交叉学科的地位,并将继续在理论、设计和应用上逐步深入。

人一出生,大脑就有精巧的构造和具有通过我们通常称为"经验"而建立它自己规则的能力。事实上经验是经时间积累的,大脑在出生后头两年内会发生非常显著的发展(即硬件线路),超越这个阶段后发展将继续进行。

一个"发展中"的大脑是与可塑的神经元同时的。可塑性允许一个发展中的神经系统适应它的周边环境。正如可塑性似乎是人类大脑中作为信息处理单元的神经元的功能的关键,它在人工神经元组成的神经网络中亦是如此。最普通的形式,神经网络是一个设计出来以对人脑完成特定任务或感兴趣功能的方法进行建模的机器;网络一般用电子器件实现或者用软件在数字计算机上模拟。在本章中,我们的兴趣大致锁定在一类重要的神经网络上,这种网络通过学习过程来实现有用的计算。为了获得好的结果,神经网络使用了一个很庞大的简单计算单元间的内部连接,这些简单计算单元称为"神经元"或者"处理单元"。据此我们给出将神经网络看作一种自适应机器的定义。

一个神经网络是一个由简单处理单元构成的规模宏大的并行分布处理器。天然具有存储经验知识和使之可用的特性。神经网络从两个方面模拟大脑。

（1）神经网络获取的知识是从外界环境中学习得来的。

（2）内部神经元的连接强度，即突触权值，用于储存获取的知识。

用于完成学习过程的程序称为学习算法，其功能是以有序的方式改变系统权值以获得想要的设计目标。

突触权值修改提供神经网络设计的传统方法。这种方法和线性自适应滤波器理论很接近，滤波器理论已经很好地建立起来并成功应用在很多不同领域。但是神经网络修改它的拓扑结构亦是可能的，这也和人的神经元会死亡和新的突触连接会建立等情况相适应。

神经网络在文献中也称为神经计算机、连接主义网络、并行分布处理器等。很明显神经网络的计算能力来自以下几点：①大规模并行分布式结构。②神经网络学习能力从而具有泛化能力。泛化是指神经网络对不在训练（学习）集中的数据可以产生合理的输出。这两个信息处理能力让神经网络可以解决一些复杂（大规模）的现在还不能处理的问题。但是实践中，神经网络不能单独作出解答，它们需要被整合在一个一致的系统工程方法中。具体地，一个感兴趣的复杂问题被分解成几个相对简单的子问题，神经网络分配与它的固有能力相符的子问题。但是，我们在建立一个可以模拟人脑的计算机结构（如果可能）之前还有很长路要走，认识这一点是很重要的。

神经网络的应用提供了下列有用的性质和能力。

（1）非线性。一个人工神经元可以是线性或者是非线性的。一个由非线性神经元互联而成的神经网络自身是非线性的，并且非线性是一种分布于整个网络中的特殊性质。非线性是一个很重要的性质，特别地，如果产生输入信号（如语音信号）内部的物理机制是天生非线性的。

（2）输入输出映射。监督学习或有教师学习是一个学习的流行的变化例，涉及使用带标号的训练样本或任务例子对神经网络的突触权值进行修改。每个样本由一个唯一的输入信号和相应期望反应组成。从一个训练集中随机选取一个样本给网络，网络就调整它的突触权值（自由参数），以最小化期望反应和由输入信号以适当的统计准则产生的实际反应之间的差别。使用训练集中的很多样本重复神经网络的训练，直到网络到达没有大的突触权值修正的稳定状态为止。先前用过的样本可能还要在训练期间以不同顺序重复使用。因此对当前问题网络通过建立输入输出映射从样本中进行学习。这样一个方法使人想起了无参数统计推断的研究，它是非模型估计的统计处理的一个分支，或者从生物学角度看，称为 tabula rasa 学习。这里使用“非参数”表示的一个事实是，没有对输入数据统计模型作任何先验假设。比如，考虑一个模式分类任务，其要求是把代表具体物体或事件的输入信号分类到几个预先分好的类中去。在这个问题的非参数方法中，要求利用例子集“估计”输入信号空间中模式分类任务的任意判决边界，并且不使用概率分布模型。监督学习隐含了一个类似的观点，神经网络的输入输出之间的映射和非参数统计推断相类似。

（3）适应性。神经网络具有调整自身连接权值以适应外界变化的能力。特别是在特定环境下可以运行的神经网络对环境的小变化可以很容易重新训练。更进一步，当它在一个时变环境（即它的统计特性随时间变化）中运行时，网络突触权值就可以设计成随时间变化。用于模式识别、信号处理、控制的神经网络耦合它的自适应能力就可以变成能进行自适应模式识别、自适应信号处理、自适应控制的有效工具。作为一个普通规则，在保证系统保持稳定时一个系统的自适应性越好，当要求在一个时变环境下运行时它就越具鲁棒性。但是，需

要强调的是,适应性不一定导致鲁棒性,实际可能相反。例如,一个暂态适应系统可能变化过快,以至于对尖峰干扰有反应,这将引起系统性能的急剧恶化。为最大限度实现适应性,系统的主要稳定时间长度应该长到可以忽略尖峰干扰,而短到可以反应环境的大体变化。这是一个稳定性和可塑性之间的二难问题[292]。

(4) 证据响应。在模式识别的问题中,神经网络可以设计成既提供选择哪一个特定模式的信息,也提供做出判决的置信度的信息。后者可以用来拒判那些出现的过于模糊的模式。有这些信息,网络的输出结果就会有改善。

(5) 前后关系的信息。神经网络的特定结构和活动状态代表知识。网络中每一个神经元都潜在地受到网络中所有其他神经元全局活动的影响。必然地,前后关系信息自然由一个神经网络处理。

(6) 容错性。一个以硬件形式实现的神经网络有天生容错的潜质,或者鲁棒计算的能力,意即它的性能在不利运行条件下逐渐下降。比如,一个神经元或一个连接损坏了,已存的模式的回忆在质量上被削弱。但是,由于网络信息存储的分布特性,在网络的总体响应严重劣化之前这种损坏是分散的。因此,原则上,一个神经网络的性能显示了一个缓慢变差的过程而不是灾难性的失败。有一些关于鲁棒性计算的经验证据,但通常它是不可控的。为了确保网络事实上的容错性,有必要在设计训练网络的算法时采用正确的度量[365]。

(7) VLSI 实现。神经网络的大规模并行性使它具有快速处理某些任务的潜在能力。这一特性使得神经网络很适合用超大规模集成电路(VLSI)实现。VLSI 的一个特殊优点是提供一种以高度分层的方式匹配真实复杂性行为的方法。

(8) 分析和设计的一致性。基本上,神经网络作为信息处理器具有通用性。我们这样说是在这样的意义下,即涉及神经网络的应用的所有领域都使用同样记号。这种特征体现在以下 3 个方面:①神经元不管形式如何,在所有的神经网络中都代表一个相同成分;②这种共性使得在不同应用中的神经网络共享相同的理论和算法成为可能;③模块化网络可以用模块的无缝整合来实现。

(9) 神经生物类比。神经网络的设计是由对大脑的类比引发的,而大脑是一个容错的并行处理的活生生的例子;神经网络的设计不仅在物理上是可实现的而且是快速高效的。神经生物学家将(人工)神经网络看作是一个解释神经生物现象的研究工具。另外,工程师关注神经生物学是将其作为解决复杂问题的新思路,这些问题比基于常规的硬件线路设计技术所能解决的问题更复杂。下面两个例子说明了这两种观点。

在文献[22]中比较了前庭视觉的线性系统模型。前庭视觉反射作用(Vestibulo-Ocular Reflex,VOR)是眼球运动系统的一部分,其作用是让眼球向与头转动方向相反的方向运动,以维持视觉(视网膜)图像的稳定性。VOR 由前庭核酸的前端神经元调节,前端神经元从前庭感知神经元中接受头部旋转信息并处理,将结果告知眼球肌肉的动作神经元。输入(头部旋转信息)和输出(眼球旋转)可以精确确定,因此 VOR 很适合用来建模。另外,它是比较简单的反射作用,并且其组成神经元的神经生理学的内容已经被很好阐述。在这三种神经模型中,前端神经元(反射内层神经元)在前庭神经核酸中是最复杂也是最引人注意的。VOR 已经用集块线性系统描述器和控制理论模型化了。这些模型对解释 VOR 的整体性质有一些作用,但是对其组成神经元特性的了解却用处不大。这种情况通过神经网络的模型已经被大大改善了。VOR 的递归网络模型能重现和解释调节 VOR 的神经元的处理信号时的静态、动态、非线性和分布性许多方面的特性,特别是前庭核酸神经元。

视网膜是我们开始将外部环境的物理图像投射到一行接受器上形成的视觉表示和第一个神经图像结合的地方。它是眼球后部的神经组织薄层，其功能是将光学图像转换成神经图像并沿光神经传输给大量的视觉中枢以便进一步处理。这是一个复杂的工作，可以从视网膜的复杂组织得到证明。在脊椎动物的视网膜中光图像转化成神经图像的过程由以下三个阶段组成[739]。

(1) 感应神经元层的图像传导。

(2) 将结果信号(对光刺激的反应产生)以化学性突触的形式传输给一层二极性单元。

(3) 同样，以化学性突触的形式把结果信号传给称为神经节的输出神经元。

在两个感应阶段(即从接受器到二极单元和从二极单元到神经节单元)，有专门侧向连接的分别称为水平连接神经元和无长突细胞的神经元。这些神经元的工作是修改感应层之间的传输。另外有被叫做中间网状细胞的分配元素，它们的工作是将信号从内部感应层传到外部感应层。一些研究人员已经建立了模拟视网膜结构的电子芯片[86]，这些电子芯片称为神经形态集成电路。一个神经形态的图像传感器由成排的感光器构成，由每个元素(像素)模拟回路结合而成。它能模拟视网膜适应局部的亮度变化、检测边缘和检测运动。神经生物学模拟，例如神经形态集成电路，有另一个重要的应用：它提供了一种希望和信念，并在一定程度上提供了一种存在性证明，即对神经生物学结构的物理上的了解对电子学工艺和超大规模集成电路技术有创造性的影响。

迄今为止，按生物神经网络大规模并行分布方式构造的各种人工神经网络，虽然已经在信息处理中扮演着越来越重要的角色，但是并没有显示出人们所期望的聪明智慧来。对以仿效大脑神经系统为目的的人工神经网络的研究历程进行一些分析和反思，探讨下一步可能采取的方法步骤，对今后智能信息科学的进一步发展将是有益的。要让人工神经网络更好更快地向人脑神经网络学习，就有必要对今后人工神经网络的主攻方向、研究路线、方法步骤、关键技术和应当采取的措施等方面，作一些考虑和调整。

概括起来，为了使人工神经网络向人脑神经网络学习得更快更好，可以从以下 3 个方面着手。

(1) 明确人工神经网络的主要智能优势是擅长非精确性信息处理。按照输入和输出特性的不同，可将智能系统最经常处理的信息类型归纳为四类。其中，数值计算和逻辑推理所对应的信息处理任务，现行的冯·诺依曼数字计算机有着成熟而巨大的能力，人工神经网络既无必要也无优势在此等领域中与之竞争。其他三类都在输入信息或输出信息中包含有非精确信息处理的内容，因而它们是人工神经网络能发挥作用的领域。人类的绝大部分脑力劳动正是投入在这 3 类信息处理上。

(2) 探索新的人工神经网络体系结构。神经解剖学方面的研究表明，在人脑神经网络中除了有由神经细胞体-轴突-突触-树突构成的神经电位脉冲电路系统外，还有一个在前后突触间约 0.02$\mu$s 的间隙中释放化学性神经递质(谷氨酸)的调节系统，这个化学递质系统的作用相当于半导体的栅极电路，起着至关重要的调控放大作用。因此，如何建立神经电脉冲系统和化学递质系统合成的耦合系统数学模型，分析其工作机理并提出可行简便算法，是一条从建模上使人工神经网络更靠近人脑神经网络的值得重视的途径。

(3) 寻求新的网络拓扑结构和相应的学习算法。迄今为止，以和积式神经元为基础的前馈型多层(特别是三层)神经网络，与误差反向传播(Back-Propagation，BP)学习算法相配合的人工神经网络结构体制，获得了最广泛的应用。但是，它在性能上仍存在若干待改进的

地方。例如,它只能调整权值,不能调整网络拓扑结构,无法实现注意力集中功能;学习新样本时,会“冲乱”原已学好保存下来的旧样本;其学习算法中,包含有较复杂的非线性激活函数的求导运算等。

## 3.2 神经元模型

从神经元的特性和功能可以知道,神经元是一个多输入单输出的信息处理单元,而且,它对信息的处理是非线性的。根据神经元的特性和功能,可以把神经元抽象为一个简单的数学模型。工程上用的人工神经元模型如图 3.1 所示。

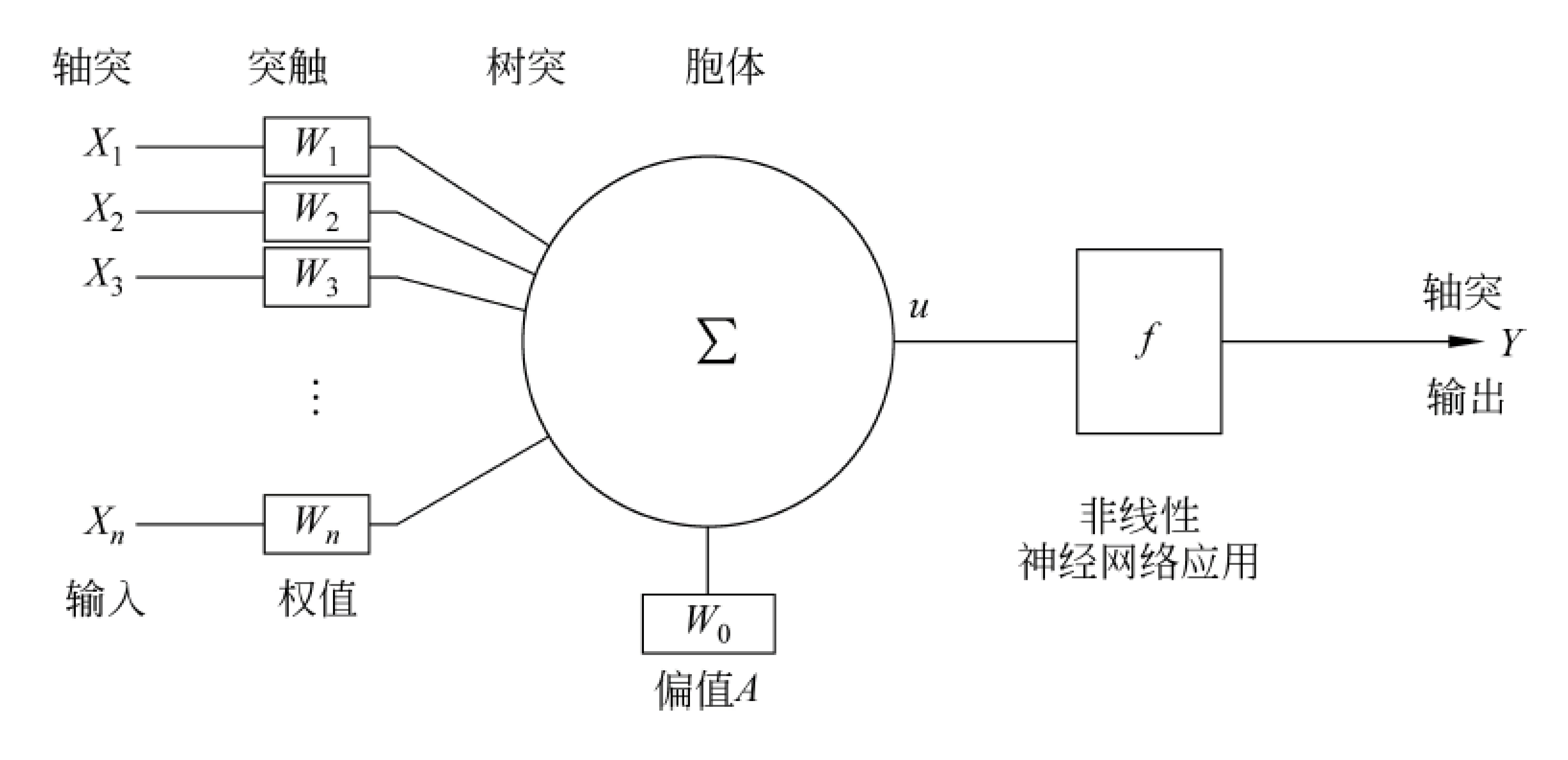



图 3.1 神经元的数学模型

在图 3.1 中,$X_1,X_2,\cdots,X_n$ 是神经元的输入,即是来自前级 $n$ 个神经元的轴突的信息。$A$ 是神经元 $i$ 的阈值;$W_1,W_2,\cdots,W_n$ 分别是神经元对 $X_1,X_2,\cdots,X_n$ 的权系数,也即突触的传递效率;$Y$ 是神经元的输出;$f$ 是激活函数,它决定神经元受到输入 $X_1,X_2,\cdots,X_n$ 的共同刺激达到阈值时以何种方式输出。

从图 3.1 的神经元模型,可以得到神经元的数学模型表达式:

$$f(u_i)=\begin{cases}1 & u_i>0\\ 0 & u_i\leqslant 0\end{cases}$$

对于激活函数 $f$ 有多种形式,其中最常见的有阶跃型、线性型和 S 形三种形式,这三种形式如图 3.2 所示。

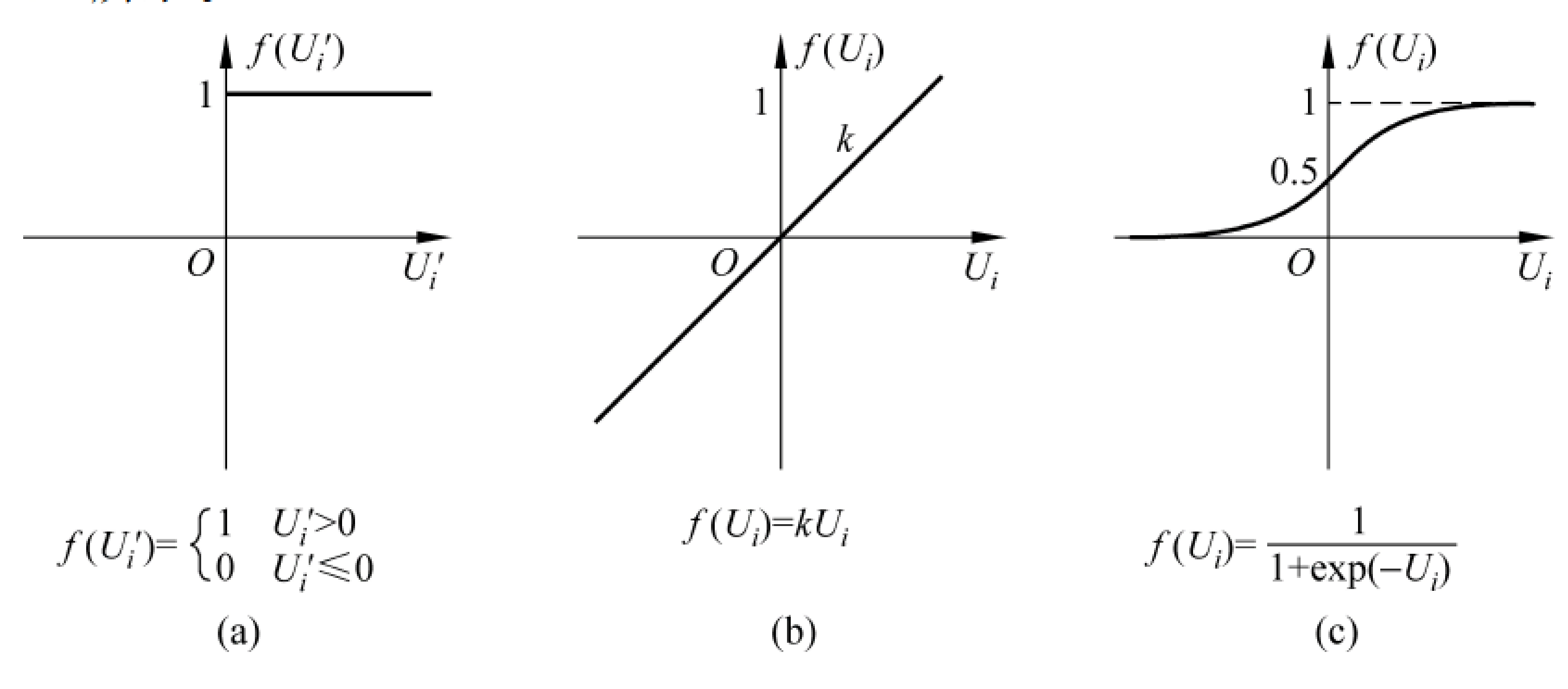



图 3.2 典型激活函数

上面所叙述的是应用最广泛而且人们最熟悉的神经元数学模型，也是历史最悠久的神经元模型。近年来，随着神经网络理论的发展，出现了不少新颖的神经元数学模型，包括逻辑神经元模型、模糊神经元模型等，也渐渐受到人们的关注和重视。

# 3.3 反传学习算法

近年来人工神经网络在很多领域得到了应用，其中大部分采用前馈网络（Feedforward）和反传算法（Backpropagation：BP）。BP 算法是为了解决多层前馈神经网络的权系数优化而提出来的，所以，BP 算法也通常暗示着神经网络的拓扑结构是一种无反馈的多层前馈网络。故而，有时也称无反馈多层前馈网络为 BP 模型。

## 3.3.1 反传算法的原理

反传算法是用于前馈多层网络的学习算法，前馈多层网络的结构一般如图 3.3 所示。它含有输入层、输出层以及处于输入输出层之间的隐层。在隐层中的神经元也称隐藏单元。隐层虽然和外界不连接，但是，它们的状态则影响输入输出之间的关系。这也是说，改变隐层的权系数，可以改变整个多层神经网络的性能。有实验表示，增加隐层的层数和隐藏单元的个数不一定能够提高网络精度和表达能力。所以 BP 网一般都选用二级网络。

反传算法分两步进行，即正向传播和反向传播。这两个过程的工作简述如下。

（1）正向传播

输入的样本从输入层经过隐藏单元一层一层进行处理，通过所有的隐层之后，则传向输出层；在逐层处理的过程中，每一层神经元的状态只对下一层神经元的状态产生影响。在输出层把当前输出和期望输出进行比较，如果当前输出不等于期望输出，则进入反向传播过程。

（2）反向传播

反向传播时，把误差信号按原来正向传播的通路反向传回，并对每个隐层的各个神经元的权系数进行修改，以期望误差信号趋向最小。

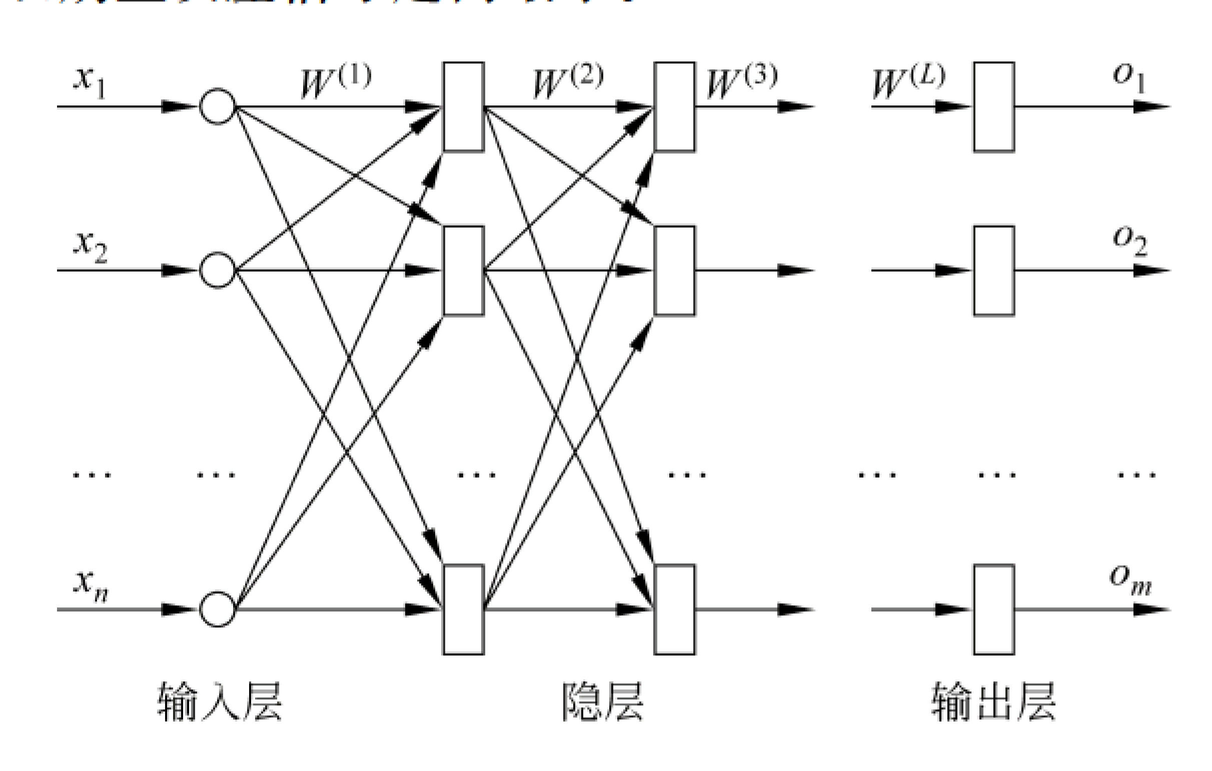



图 3.3 前馈多层网络的结构

设有一个 $m$ 层的神经网络，并在输入层加有样本 $X$；设第 $k$ 层的 $i$ 神经元的输入总和表示为 $U_i^k$，输出 $X_i^k$；从第 $k-1$ 层的第 $j$ 个神经元到第 $k$ 层的第 $i$ 个神经元的权系数为 $W_{ij}$，各个神经元的激发函数为 $f$，则各个变量的关系可用下面有关数学式表示：

$$X_i^k = f(U_i^k) \tag{3.1}$$

$$U_i^k = \sum_j W_{ij} X_j^{k-1} \tag{3.2}$$

## 3.3.2 反传算法的数学表达

反传算法实质是求取误差函数的最小值问题。这种算法采用非线性规划中的最速下降方法,按误差函数的负梯度方向修改权系数。为了说明反传算法,首先定义误差函数 $e$。取期望输出和实际输出之差的平方和为误差函数,则有

$$e = \frac{1}{2}\sum_i (X_i^m - Y_i)^2 \tag{3.3}$$

其中:$Y_i$ 是输出单元的期望值,它也在这里用作教师信号;$X_i^m$ 是实际输出;因为第 $m$ 层是输出层。

由于反传算法按误差函数 $e$ 的负梯度方向修改权系数,故权系数 $W_{ij}$ 的修改量 $\Delta W_{ij}$

$$\Delta W_{ij} \infty - \frac{\partial e}{\partial W_{ij}}$$

也可写成

$$\Delta W_{ij} \infty - \eta \frac{\partial e}{\partial W_{ij}}$$

其中:$\eta$ 为学习速率,即步长,一般取 0~1 的数。

很明显,根据 BP 算法原则,求$\frac{\partial e}{\partial W_{ij}}$最关键的。下面求$\frac{\partial e}{\partial W_{ij}}$有:

$$\frac{\partial e}{\partial W_{ij}} = \frac{\partial e_k}{\partial U_i^k} \cdot \frac{\partial U_i^k}{\partial W_{ij}} \tag{3.4}$$

由于 $\frac{\partial U_i^k}{\partial W_{ij}} = \frac{\partial\left(\sum_l W_{il} X_l^{k-l}\right)}{\partial W_{ij}} = X_j^{k-l} \mid_{l=j}$ 故

$$\frac{\partial e}{\partial W_{ij}} = \frac{\partial e_k}{\partial U_i^k} \cdot X_j^{k-l} \tag{3.5}$$

从而有

$$\Delta W_{ij} = -\eta \frac{\partial e}{\partial W_{ij}} = -\eta \frac{\partial e_k}{\partial U_i^k} \cdot X_j^{k-l} \tag{3.6}$$

令

$$d_i^k = \frac{\partial U_i^k}{\partial W_{ij}} \tag{3.7}$$

则有学习公式:

$$\Delta W_{ij} = -\eta d_i^k \cdot X_j^{k-l} \tag{3.8}$$

其中:$\eta$ 为学习速率,即步长,一般取 0~1 的数。

上面实际仍未给出求 $d_i^k$ 明显的算法公式,求 $d_i^k$ 的计算公式如下:

$$d_i^k = \frac{\partial U_i^k}{\partial W_{ij}} = \frac{\partial e}{\partial X_i^k} \cdot \frac{\partial X_i^k}{\partial U_i^k} \tag{3.9}$$

由于

$$\frac{\partial X_i^k}{\partial U_i^k} = f'(U_i^k)$$

为了方便进行求导，取 $f$ 为连续函数。一般取非线性连续函数，例如 Sigmoid 函数。当取 $f$ 为非对称 Sigmoid 函数时，有

$$f(U_i^k) = \frac{1}{1+\exp(-U_i^k)} \tag{3.10}$$

则有

$$f'(U_i^k) = f'(U_i^k)(1-f(U_i^k)) = X_i^k(1-X_i^k)$$

再考虑式(3.9)中的偏微分项 $\frac{\partial e}{\partial X_i^k}$，有以下两种情况：

(1) 如果 $k=m$，则是输出层，这时有 $Y_i$ 是输出期望值，它是常数。从式(3.3)有

$$\frac{\partial e}{\partial X_i^k} = \frac{\partial e}{\partial X_i^m} = (X_i^m - Y_i) \tag{3.11}$$

从而有

$$d_i^m = X_i^m(1-X_i^m)(X_i^m-Y_i) \tag{3.12}$$

(2) 如果 $k<m$，则该层是隐层。这时应考虑上一层对它的作用，故有

$$\frac{\partial e}{\partial X_i^k} = \sum_l \frac{\partial e}{\partial U_l^{k+1}} \cdot \frac{\partial U_l^{k+1}}{\partial X_i^k} \tag{3.13}$$

由式(3.7)可知

$$\frac{\partial e}{\partial U_l^{k+1}} = d_l^{k+1} \tag{3.14}$$

由式(3.2)可知

$$\frac{\partial U_i^k}{\partial X_i^k} = \frac{\partial\left(\sum_j W_{ij}X_j^{k-1}\right)}{\partial X_i^k} = W_{li}\ |_{j=i} \tag{3.15}$$

故有

$$\frac{\partial e}{\partial X_i^k} = \sum_l W_{li} \cdot d_l^{k+l} \tag{3.16}$$

最后有

$$d_i^k = X_i^k(1-X_i^k) \cdot \sum_l W_{li} \cdot d_l^{k+l} \tag{3.17}$$

从上述过程可知：多层网络的训练方法是把一个样本加到输入层，并根据向前传播的规则 $X_i^k=f(U_i^k)$，不断一层一层向输出层传递，最终在输出层可以得到输出 $X_i^m$。

把 $X_i^m$ 和期望输出 $Y_i$ 进行比较。如果两者不等，则产生误差信号 $e$，接着则按下面公式反向传播修改权系数：

$$\Delta W_{ij} = -\eta d_i^k \cdot X_j^{k-l} U_i^k = \sum_j W_{ij} X_j^{k-1} \tag{3.18}$$

其中，$d_i^m=X_i^m(1-X_i^m)(X_i^m-Y_i)$，$d_i^k = X_i^k(1-X_i^k) \cdot \sum_l W_{li} \cdot d_l^{k+l}$。

式(3.18)中，求取本层 $d_i^k$ 时，要用到高一层的 $d_i^k+1$；可见，误差函数的求取是从输出层开始，到输入层的反向传播过程。在这个过程中不断进行递归求误差。

通过多个样本的反复训练，同时向误差逐渐减小的方向对权系数进行修正，最终减少误

差。从式(3.18)也可以知道,当网络的层数较多时,所需的计算量很大,导致收敛速度不快。

为了加快收敛速度,一般以上一次的权系数作为本次修正的依据之一,故而有修正公式:

$$\Delta W_{ij}(t+l) = -\eta d_i^k \cdot X_j^{k-l} + \alpha \Delta W_{ij}(t) \tag{3.19}$$

其中,$\eta$ 为学习速率,即步长,$\eta$ 取 0.1~0.4,$\alpha$ 为权系数修正常数,取 0.7~0.9。

上面式(3.18)也称为通用的 Delta 法则。对于没有隐层的神经网络,可取

$$\Delta W_{ij} = \eta(Y_j - X_j) \cdot X_i \tag{3.20}$$

其中,$Y_j$ 为期望输出;$X_j$ 为输出层的实际输出;$X_i$ 为输入层的输入。这显然是一种十分简单的情况,式(3.20)也称为简单 Delta 法则。

在实际应用中,只有通用的 Delta 法则式(3.18)或式(3.19)才有意义。简单 Delta 法则式(3.20)只在理论推导上有用。

### 3.3.3 反传算法的执行步骤

在把反传算法应用于前馈多层网络,采用 Sigmoid 为激活函数时,可用下列步骤对网络的权系数 $W_{ij}$ 进行递归求取。注意对于每层有 $n$ 个神经元的时候,即有 $i=1,2,\cdots,n$;$j=1,2,\cdots,n$。对于第 $k$ 层的第 $i$ 个神经元,则有 $n$ 个权系数 $W_{i1},W_{i2},\cdots,W_{in}$,另外再取一个 $W_{in+1}$ 用于表示阈值 $\theta_i$;并且在输入样本 $X$ 时,取 $X=(X_1,X_2,\cdots,X_n,1)$。

算法的执行的步骤如下。

(1) 对权系数 $W_{ij}$ 置初值。

对各层的权系数 $W_{ij}$ 置一个较小的非零随机数,但其中 $W_{i,n+1}=-\theta$。

(2) 输入一个样本 $X=(X_1,X_2,\cdots,X_n,1)$,以及对应期望输出 $Y=(Y_1,Y_2,\cdots,Y_n)$。

(3) 计算各层的输出。

对于第 $k$ 层第 $i$ 个神经元的输出 $X_i^k$,有:

$$U_i^k = \sum_{j=1}^{n+1} W_{ij} X_j^{k-1},\quad X_{n+1}^{k-1} = 1,\quad W_{i,n+1} = -\theta,\quad X_i^k = f(U_i^k)$$

(4) 求各层的学习误差 $d_i^k$。

对于输出层 $k=m$,有

$$d_i^m = X_i^m(1 - X_i^m)(X_i^m - Y_i)$$

对于其他各层,有

$$d_i^k = X_i^k(1 - X_i^k) \cdot \sum_l W_{li} \cdot d_l^{k+l}$$

(5) 修正权系数 $W_{ij}$ 和阈值 $\theta$,用式(3.18)时有:

$$\Delta W_{ij}(t+l) = \Delta W_{ij}(t) - \eta \cdot d_i^k \cdot X_j^{k-1}$$

用式(3.19)时有:

$$\Delta W_{ij}(t+l) = \Delta W_{ij}(t) - \eta \cdot d_i^k \cdot X_j^{k-1} + \alpha \Delta W_{ij}(t)$$

其中,

$$\Delta W_{ij}(t) = -\eta \cdot d_i^k \cdot X_j^{k-1} + \alpha \Delta W_{ij}(t-l) = W_{ij}(t) - W_{ij}(t-l)$$

(6) 当求出了各层各个权系数之后,可按给定品质指标判别是否满足要求。如果满足要求,则算法结束;如果未满足要求,则返回(3)执行。

这个学习过程，对于任一给定的样本 $X_p=(X_{p1},X_{p2},\cdots,X_{pn},1)$ 和期望输出 $Y_p=(Y_{p1},Y_{p2},\cdots,Y_{pn})$ 都要执行，直到满足所有输入输出要求为止。

## 3.3.4　对反传网络优缺点的讨论

多层前向反传网络是目前应用最广泛的一种神经网络模型，但它也不是尽善尽美的，为了更好地理解应用神经网络进行问题求解，这里对它的优缺点展开讨论。

### 1. 多层前向反传网络的优点

(1) 网络实质上实现了一个从输入到输出的映射功能，而数学理论已证明它具有实现任何复杂非线性映射的功能。因此它特别适合于求解内部机制复杂的问题。

(2) 网络能通过学习带正确答案的实例集，自动提取“合理的”求解规则，即具有自学习能力。

(3) 网络具有一定的泛化、概括能力。

### 2. 多层前向反传网络的不足

(1) BP 算法的学习速度很慢，其原因主要有：①由于 BP 算法本质上为梯度下降法，而它所要优化的目标函数又非常复杂，因此，必然会出现“锯齿形现象”，这使得 BP 算法效率下降；②存在麻痹现象，由于优化的目标函数很复杂，它必然会在神经元输出接近 0 或 1 的情况下，出现一些平坦区，在这些区域内，权值误差改变很小，使训练过程几乎停顿；③为了使网络执行 BP 算法，不能用传统的一维搜索法求每次迭代的步长，而必须把步长的更新规则预先赋予网络，这种方法将引起算法低效。

(2) 网络训练失败的可能性较大，其原因主要有：①从数学角度看，BP 算法为一种局部搜索的优化方法，但它要解决的问题为求解复杂非线性函数的全局极值，因此，算法很有可能陷入局部极值，使训练失败；②网络的逼近、泛化能力同学习样本的典型性密切相关，而从问题中选取典型样本实例组成训练集是一个很困难的问题。

(3) 难以解决应用问题的实例规模和网络规模间的矛盾。这涉及网络容量的可能性与可行性的关系问题，即学习复杂性问题。

(4) 网络结构的选择尚无一种统一而完整的理论指导，一般只能由经验选定。为此，有人称神经网络的结构选择为一种艺术。而网络的结构直接影响网络的逼近能力及泛化性质。因此，应用中如何选择合适的网络结构是一个重要的问题。

(5) 新加入的样本要影响已学习成功的网络，而且刻画每个输入样本的特征的数目也必须相同。

(6) 网络的预测能力(也称泛化能力、推广能力)与训练能力(也称逼近能力、学习能力)间的矛盾。一般情况下，训练能力差时，预测能力也差，并且一定程度上，随着训练能力的提高，预测能力也提高。但这种趋势有一个极限，当达到此极限时，随训练能力的提高，预测能力反而下降，即出现所谓“过拟合”现象。此时，网络学习了过多的样本细节，而不能反映样本内含的规律。

# 3.4 Hopfield 模型

1982 年,霍普菲尔德提出了可用作联想存储器的互连网络,这个网络称为 Hopfield 网络模型,也称 Hopfield 模型。Hopfield 网络模型是一种循环神经网络,从输出到输入有反馈连接。Hopfield 网络有离散型和连续型两种。

反馈神经网络由于其输出端又反馈到其输入端,Hopfield 网络在输入的激励下,会产生不断的状态变化。当有输入之后,可以求取出 Hopfield 的输出,这个输出反馈到输入从而产生新的输出,这个反馈过程一直进行下去。如果 Hopfield 网络是一个能收敛的稳定网络,则这个反馈与迭代的计算过程所产生的变化越来越小,一旦到达了稳定平衡状态,Hopfield 网络就会输出一个稳定的恒值。对于一个 Hopfield 网络来说,关键在于确定它在稳定条件下的权系数。

应该指出:反馈网络有稳定的,也有不稳定的。对于 Hopfield 网络来说,还存在如何判别它是稳定网络还是不稳定网络,依据什么进行判别的问题。

## 3.4.1 离散 Hopfield 网络

Hopfield 最早提出的网络是二值神经网络,神经元的输出只取 1 和 0 这两个值,所以,也称为离散 Hopfield 神经网络。在离散 HopfieId 网络中,所采用的神经元是二值神经元;故而,所输出的离散值 1 和 0 分别表示神经元处于激活和抑制状态。

首先考虑由三个神经元组成的离散 Hopfield 神经网络,其结构如图 3.4 所示。

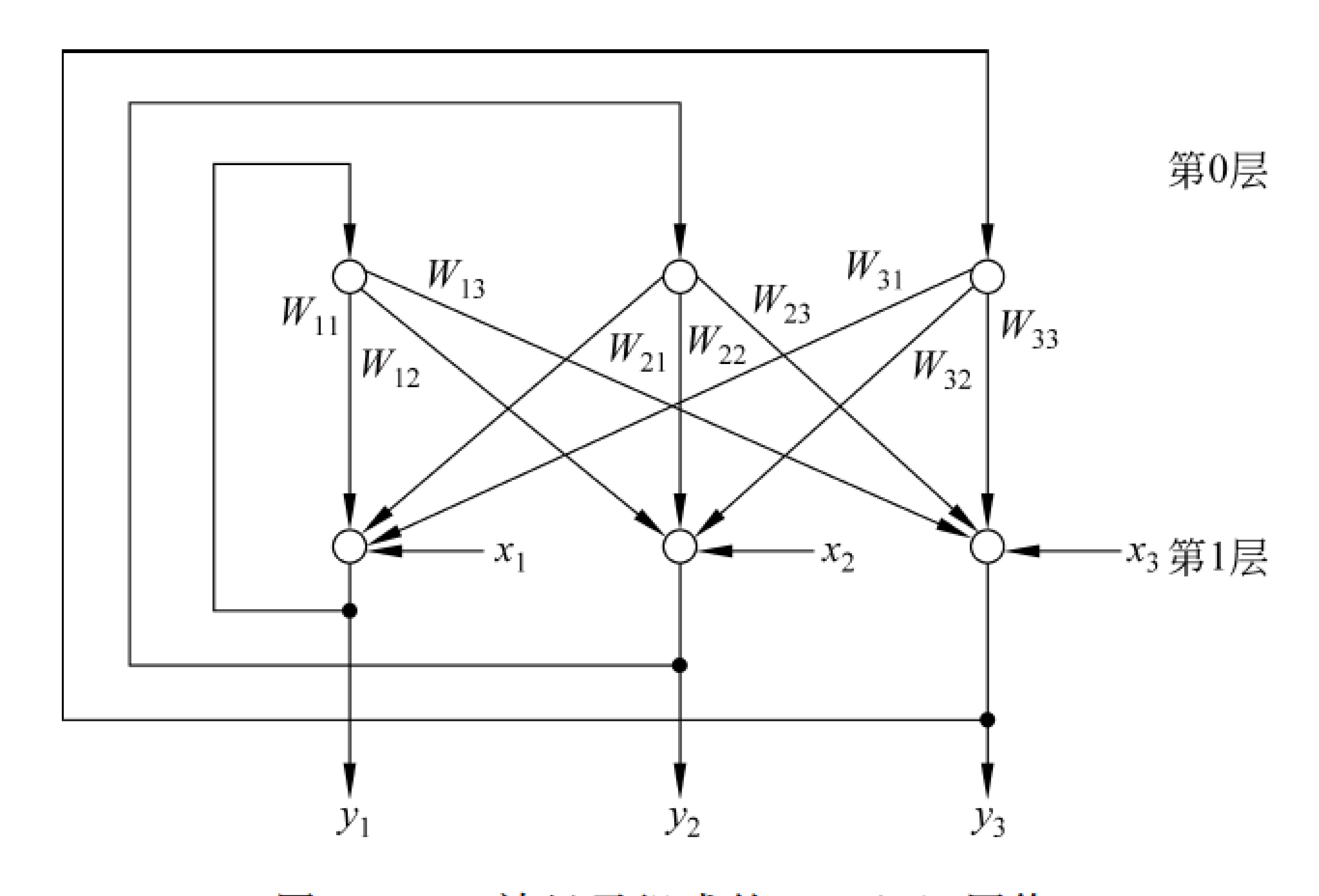



图 3.4 三神经元组成的 Hopfield 网络

在图 3.4 中,第 0 层仅仅是作为网络的输入,它不是实际神经元,所以无计算功能;而第一层是实际神经元,故而对输入信息和权系数乘积求累加和,并由非线性函数 $f$ 处理后产生输出信息。$f$ 是一个简单的阈值函效,如果神经元的输出信息大于阈值 $\theta$,那么,神经元的输出就取值为 1;小于阈值 $\theta$,则神经元的输出就取值为 0。

对于二值神经元,它的计算公式如下:

$$U_j = \sum_i W_{ij} Y_i + X_j$$

其中，$X_j$ 为外部输入。并且有：

$$Y_i = \begin{cases} 1, & U_i \geqslant \theta_i \\ 0, & U_i < \theta_i \end{cases}$$

对于一个离散的 Hopfield 网络，其网络状态是输出神经元信息的集合。对于一个输出层是 $n$ 个神经元的网络，则其 $t$ 时刻的状态为一个 $n$ 维向量：

$$\boldsymbol{Y}(t) = [Y_1(t), Y_2(t), \cdots, Y_n(t)]^{\mathrm{T}}$$

故而，网络有 $2^n$ 个状态；因为 $Y_j(t)(j=1,\cdots,n)$ 可以取值为 1 或 0；故 $n$ 维向量 $\boldsymbol{Y}(t)$ 有 $2^n$ 种状态，即网络状态。

对于三个神经元的离散 Hopfield 网络，它的输出层就是 3 位二进制数；每一个 3 位二进制数就是一种网络状态，从而共有 8 个网络状态。这些网络状态如图 3.5 所示，图中立方体的每一个顶角表示一种网络状态。同样，对于 $n$ 个神经元的输出层，它有 $2^n$ 个网络状态，也和一个 $n$ 维超立方体的顶角相对应。

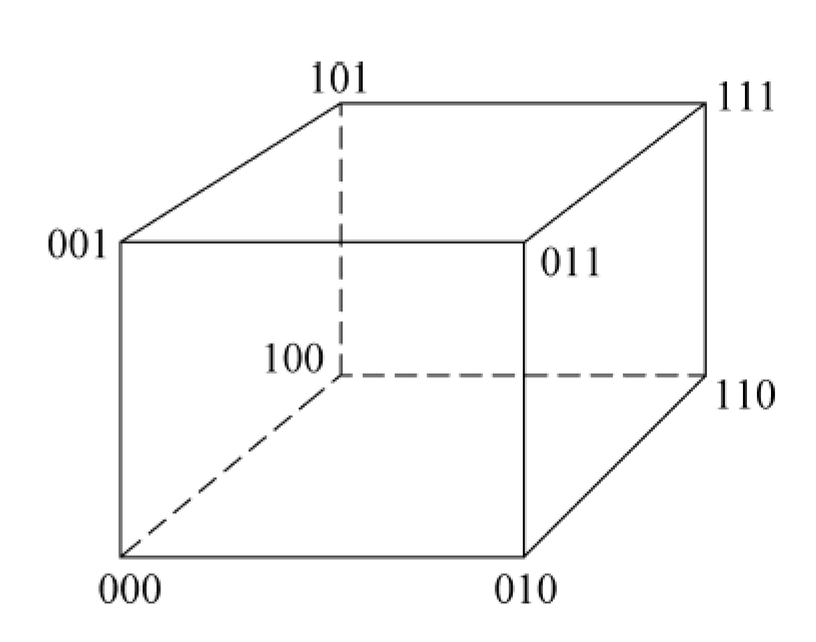



图 3.5　三神经元输出层的网络状态

如果 Hopfield 网络是一个稳定网络，那么在网络的输入端加入一个输入向量，则网络的状态会产生变化，也就是从超立方体的一个顶角转移向另一个顶角，并且最终稳定于一个特定的顶角。

对于一个由 $n$ 个神经元组成的离散 Hopfield 网络，则有 $n\times n$ 权系数矩阵 $\boldsymbol{W}$：

$$\boldsymbol{W} = \{W_{ij}\} \quad i = 1,2,\cdots,n, j = 1,2,\cdots,n$$

同时，有 $n$ 维阈值向量$\boldsymbol{\theta}$：

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_n]^{\mathrm{T}}$$

一般而言，$\boldsymbol{W}$ 和$\boldsymbol{\theta}$ 可以确定一个唯一的离散 Hopfield 网络。对于图 3.4 所示的三神经元组成的 Hopfield 网络，也可以改用图 3.6 所示的图形表示，这两个图形的意义是一样的。考虑离散 Hopfield 网络的一般节点状态；用 $Y_j(t)$表示第 $j$ 个神经元，即节点 $j$ 在时刻 $t$ 的状态，则节点的下一个时刻$(t+1)$的状态如下：

$$Y_i(t+1) = f[U_j(t)] = \begin{cases} 1, & U_j \geqslant 0 \\ 0, & U_j < 0 \end{cases} \tag{3.21}$$

$$U_j(t) = \sum_{i=1}^{n} W_{ij} Y_i(t) + X_j - \theta_j \tag{3.22}$$

当 $W_{ij}$ 在 $i=j$ 时等于 0，则说明一个神经元的输出并不会反馈到它自己的输入，这时，离散的 Hopfield 网络称为无自反馈网络；当 $W_{ij}$ 在 $i=j$ 时不等于 0，则说明一个神经元的输出会反馈到它自己的输入，这时，离散的 Hopfield 网络称为有自反馈的网络。

离散 Hopfield 网络有如下两种不同的工作方式。

### 1. 串行(异步)方式

在时刻 $t$ 时，只有某一个神经元 $j$ 的状态产生变化，而其他 $n-1$ 个神经元的状态不变，

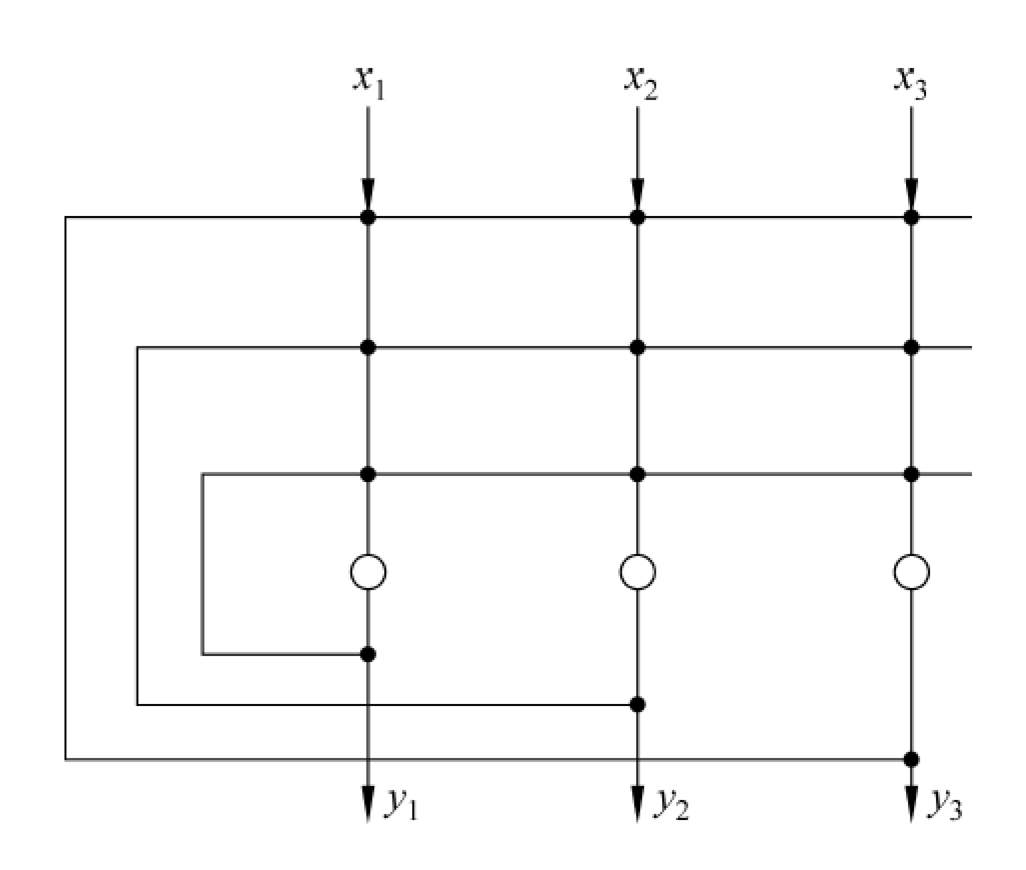



图 3.6　离散 Hopfield 网络的另外一种图示

这时称串行工作方式。并且有

$$Y_i(t+1) = f\left[\sum_{i=1}^{n} W_{ij} Y_i(t) + X_j - \theta_j\right] \tag{3.23}$$

$$Y_i(t+1) = Y_j(t) \quad i \neq j$$

在不考虑外部输入时,则有

$$Y_i(t+1) = f\left[\sum_{i=1}^{n} W_{ij} Y_i(t) - \theta_j\right]$$

### 2. 并行(同步)方式

在任一时刻 $t$,所有的神经元的状态都产生了变化,则称并行工作方式,并且有

$$Y_i(t+1) = f\left[\sum_{i=1}^{n} W_{ij} Y_i(t) + X_j - \theta_j\right] \quad j = 1,2,\cdots,n \tag{3.24}$$

在不考虑外部输入时,则有

$$Y_i(t+1) = f\left[\sum_{i=1}^{n} W_{ij} Y_i(t) - \theta_j\right] \quad j = 1,2,\cdots,n$$

对于一个网络来说,稳定性是一个重大的性能指标。对于离散 Hopfield 网络,其状态为 $Y(t)$:

$$Y(t) = [Y_1(t), Y_2(t), \cdots, Y_n(t)]^{\mathrm{T}}$$

如果,对于任何 $\Delta t > 0$。当神经网络从 $t=0$ 开始,有初始状态 $Y(0)$;经过有限时刻 $t$,有:

$$Y(t+\Delta t) = Y(t)$$

则称网络是稳定的。

在串行方式下的稳定性称之为串行稳定性。同理,在并行方式的稳定性称之为并行稳定性。在神经网络稳定时,其状态称稳定状态。

Hopfield 网络的一个功能是可用于联想记忆,也即联想存储器。这是人类的智能特点之一。人类的所谓"触景生情"就是见到一些类同过去接触的景物,容易产生对过去情景的回味和思忆。对于 Hopfield 网络,用它作联想记忆时,首先通过一个学习训练过程确定网络中的权系数,使所记忆的信息在网络的 $n$ 维超立方体的某一个顶角处的能量最小。当网

络的权系数确定之后，只要为网络给出输入向量，这个向量可能是局部数据，即不完全或部分不正确的数据，但是网络仍然产生所记忆的信息的完整输出。1984 年 Hopfield 开发了一种用 $n$ 维 Hopfield 网络作联想存储器的结构。在这个网络中，权系数的赋值规则为存储向量的外积存储规则(Out Product Storage Prescription)，其原理如下：

设有 $m$ 个样本存储向量 $X_1, X_2, \cdots, X_m$，则有

$$X_1 = \{X_{11}, X_{12}, \cdots, X_{1m}\}$$
$$X_2 = \{X_{21}, X_{22}, \cdots, X_{2m}\}$$
$$\cdots$$
$$X_m = \{X_{m1}, X_{m2}, \cdots, X_{mm}\}$$

把这 $m$ 个样本向量存储入 Hopfield 网络中，则在网络中第 $i,j$ 两个节点之间权系数的值为：

$$W_{ij} = \begin{cases} \sum X_{ik}X_{jk}, & i \neq j \\ 0, & i = j \end{cases} \tag{3.25}$$

其中：$k$ 为样本向量 $X_k$ 的下标，$k=1,2,\cdots,m$；$i,j$ 分别是样本向量 $X_k$ 的第 $i,j$ 分量 $X_i, X_j$ 的下标；$i,j=1,2,\cdots,n$。

对联想存储器的联想检索过程如下。

给定一个向量 $X$，进行联想检索求取在网络中的存储内容。这时，把向量

$$X = \{X_1, X_2, \cdots, X_n\}$$

的各个分量 $X_1, X_2, \cdots, X_n$ 赋于相对应的节点 $j$，($j=1,2,\cdots,n$)，则节点有相应的初始状态 $Y_j(0)$，则有

$$Y_j(0) = X_j, \quad j = 1,2,\cdots,n$$

接下来，在 Hopfield 网络中按动力学系统原则进行计算，得

$$Y_j(t+1) = f[\Sigma W_{ij} Y_j(0) - \theta_j], \quad i,j = 1,2,\cdots,n$$

其中，$f$ 是非线性函数，可取阶跃函数。

状态通过不断变化，最终会稳定下来.最终的状态是和给定向量 $X$ 最接近的样本向量。所以，Hopfield 网络的最终输出也就是给定向量联想检索结果。这个过程说明，即使给定向量并不完全或部分不正确，也能找到正确的结果。在本质上，它也有滤波功能。

### 3.4.2 连续 Hopfield 网络

连续 Hopfield 网络的拓扑结构和离散 Hopfield 网络相同。这种拓扑结构和生物的神经系统中大量存在的神经反馈回路是一致的。连续 Hopfield 网络和离散 Hopfield 网络一样，其稳定条件也是 $W_{ij} = W_{ji}$。连续 Hopfield 网络和离散 Hopfield 网络不同点在于其函数 $g$ 不是阶跃函数，而是 S 形的连续函数。一般取 $g(u) = 1/(1+e^{-u})$。

连续 Hopfield 网络在时间上是连续的，所以，网络中各神经元是以同步方式工作的。对于一个神经细胞，即神经元 $j$，其内部膜电位状态用 $u_j$ 表示，细胞膜输入电容为 $C_j$，细胞膜的传递电阻为 $R_j$，输出电压为 $V_j$，外部输入电流用 $I_j$ 表示，则连续 Hopfield 网络可用图 3.7 所示的电路表示。

$$\begin{cases} C_j \dfrac{\mathrm{d}U_i(t)}{\mathrm{d}t} = \sum_{j=1}^{n} W_{ij} V_j(t) - \dfrac{U_j(t)}{R_j} + I_j \\ V_j(t) = g_j(U_j(t)) \quad j = 1,2,\cdots,n \end{cases} \tag{3.26}$$

其中,$n$ 是神经网络神经元的个数,$v_j(t)$为输出电位,$U_j(t)$为输入电位。

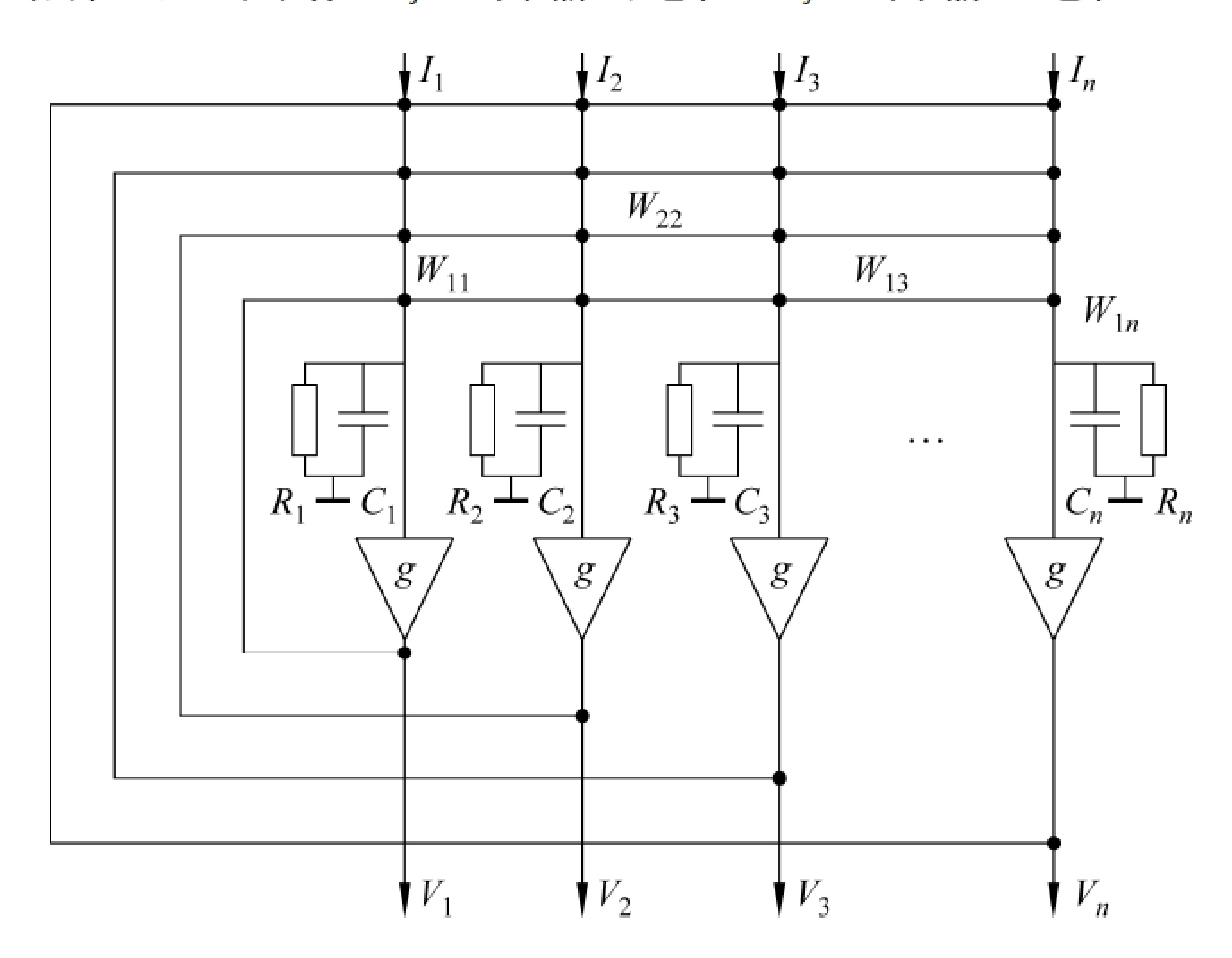



图 3.7 连续 Hopfield 网络的电路形式

对于连续 Hopfield 网络,Hopfield 给出如下稳定性定理。

给出能量函数

$$E(t) = -\frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} W_{ij} V_i(t) V_j(t) - \sum_{j=1}^{n} V_j(t) I_j + \sum_{j=1}^{n} \frac{1}{R_i}\int_0^{V_j(t)} g^{-1}(V)\mathrm{d}V \tag{3.27}$$

其中,$g^{-1}(V)$是 $V_j(t)=g_j(U_j(t))$的反函数。

如果连续 Hopfield 网络中神经元传递函数是单调增长的连续有界函数,并且 $W_{ij} = W_{ji}$,则有$\dfrac{\mathrm{d}E(t)}{\mathrm{d}t}\leqslant 0$;当且仅当$\dfrac{\mathrm{d}V_i(t)}{\mathrm{d}t}\leqslant 0$ 有$\dfrac{\mathrm{d}E(t)}{\mathrm{d}t}=0$。

这个定理的意义可以解释如下:当网络神经元的传递函数是 S 函数,并且网络权系数矩阵对称,则随时间的变化网络的能量会下降或不变;而且仅当输出电位随时间变化不变时,网络的能量才会不变。换而言之,在上述条件下的网络是能量不变或下降的。

这个定理说明 Hopfield 网络的能量函数 $E(t)$是单调下降的;如果 $E(t)$有下界,即有确定的极小值;那么网络必定是稳定的。而且,可以知道稳定点对应于能量函数的下界,即极小值。

下一步工作,只需证明能量函数有下界,那么,就可以证明网络是稳定的。

可以证明,如果 Hopfield 网络的传递函数 $g$ 是连续有界的,那么,能量函数 $E(t)$是有界的。

最后,有如下结论:当 Hopfield 网络的神经元传递函数 $g$ 是连续且有界的,例如 Sigmoid 函数,并且网络的权系数矩阵对称,则这个连续 Hopfield 网络是稳定的。在实际应用中,任何一个系统,如果其优化问题可以用能量函数 $E(t)$作为目标函数,那么,总可以用连续 Hopfield 网络对其进行求解。由于引入能量函数 $E(t)$,Hopfield 使神经网络和问题优化直接对应;这种工作是具有开拓性的。利用神经网络进行优化计算,就是在神经网络这

一动力系统给出初始的估计点，即初始条件；然后随网络的运动传递而找到相应极小点。这样，大量的优化问题都可以用连续的 Hopfield 网来求解。这也是 Hopfield 网络用于神经计算的基本原因。

## 3.5　自适应共振理论 ART 模型

自适应共振理论（Adaptive Resonance Theory，ART）模型是美国波士顿大学的格罗斯伯格（S Grossberg）在 1976 年提出的。ART 是一种自组织神经网络结构，是无监督的学习网络。当在神经网络和环境有交互作用时，对环境信息的编码会自发地在神经网中产生，则认为神经网络在进行自组织活动。ART 就是这样一种能自组织地产生对环境认识编码的神经网络理论模型。

ART 模型是基于下列问题的求解而提出的。

（1）对于一个学习系统，要求它有适应性及稳定性，适应性可以响应重要事件，稳定性可以存储重要事件。这种系统的设计问题。

（2）学习时，原有的信息和新信息如何处理，保留有用知识，接纳新知识的关系如何解决的问题。

（3）对外界信息与原来存储的信息结合并决策的问题。

格罗斯伯格一直对人类的心理和认识活动感兴趣，他长期致力于这方面的研究并希望用数学来刻画人类这项活动，建立人类的心理和认知活动的一种统一的数学模型和理论。ART 就是由这种理论的核心内容并经过提高发展然后得出的。

目前，ART 理论已提出了三种模型结构，即 ART1，ART2，ART3。ART1 用于处理二进制输入的信息；ART2 用于处理二进制和模拟信息这两种输入；ART3 用于进行分级搜索。ART 理论可以用于语音、视觉、嗅觉和字符识别等领域。

### 3.5.1　ART 模型的结构

ART 模型源于 Helmholtz 无意识推理学说的协作-竞争网络交互模型[289]。这个模型如图 3.8 所示。从图中看出这个模型由两个协作-竞争模型组成。无意识推理学说认为：原始的感觉信息通过经历过的学习过程不断修改，直到得到一个真实的感知结果为止。在图 3.8 中协作-竞争网络交互模型可以看出：环境输入信号和自上而下学习期望同时对协作-竞争网络 1 执行输入；而自下而上学习是协作-竞争网络 1 的输出；同时，自下而上学习是协作-竞争网络 2 的输入，而自上而下学习期望则是其输出。真实感知是通过这个协作-竞争网络的学习和匹配产生的。

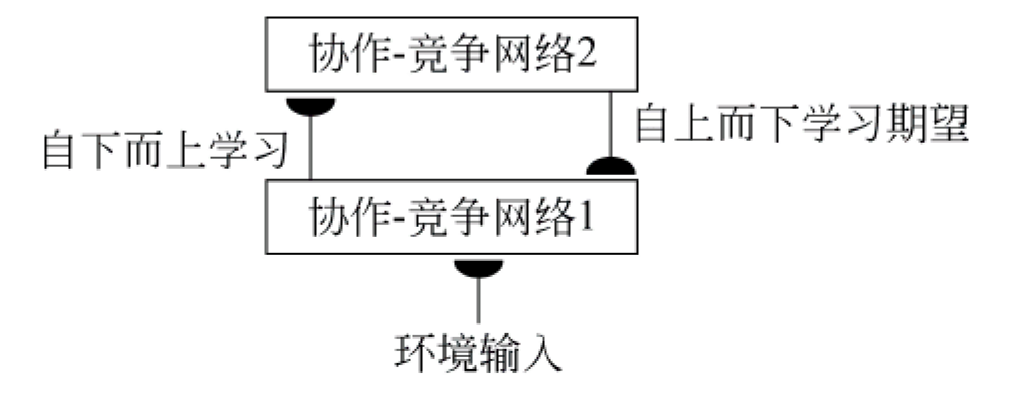



图 3.8　协作-竞争网络交互模型

环境输入信号对自上而下学习期望进行触发,使协作-竞争网络1产生自下而上学习的输出。输出发送到协作-竞争网络2,则产生自上而下学习期望输出,并送回协作-竞争网络1。这个过程很明显是自上而下学习和自下而上学习的过程,并且这个过程中不断吸收环境输入信息。经过协作-竞争的匹配,最终取得一致的结果,这也就是最终感知或谐振感知。协作-竞争网络交互作用有下列基本要求。

(1) 交互作用是非局域性的。

(2) 交互作用是非线性的。

(3) 自上而下的期望学习是非平稳随机过程。

受到协作-竞争网络交互模型的启发,格罗斯伯格提出了ART理论模型。他认为对网络的自适应行为进行分析,可以建立连续非线性网络模型,这种网络可以由短期存储STM和长期存储LTM作用所实现。STM是指神经元的激活值,即未由S函数处理的输出值,LTM是指权系数。

格罗斯伯格所提出的ART理论模型有如下一些主要优点。

(1) 可以进行实时学习,能适应非平稳的环境。

(2) 对于已经学习过的对象具有稳定的快速识别能力;同时,亦能迅速适应未学习的新对象。

(3) 具有自归一能力,根据某些特征在全体中所占的比例,有时作为关键特征,有时当作噪声处理。

(4) 不需要预先知道样本结果,是无监督学习;如果对环境作出错误反映则自动提高"警觉性",迅速识别对象。

(5) 容量不受输入通道数的限制,存储对象也不是正交的。

ART的基本结构如图3.9所示。它由输入神经元和输出神经元组成。用前向权系数和样本输入来求取神经元的输出,这个输出也就是匹配测度,具有最大匹配测度的神经元的活跃级通过输出神经元之间的横向抑制得到进一步增强,而匹配测度不是最大的神经元的活跃级就会逐渐减弱,从输出神经元到输入神经元之间有反馈连接以进行学习比较。同样,还提供一个用来确定具有最大输出的输出神经元与输入模式进行比较的机制。ART模型的框图如图3.10所示。

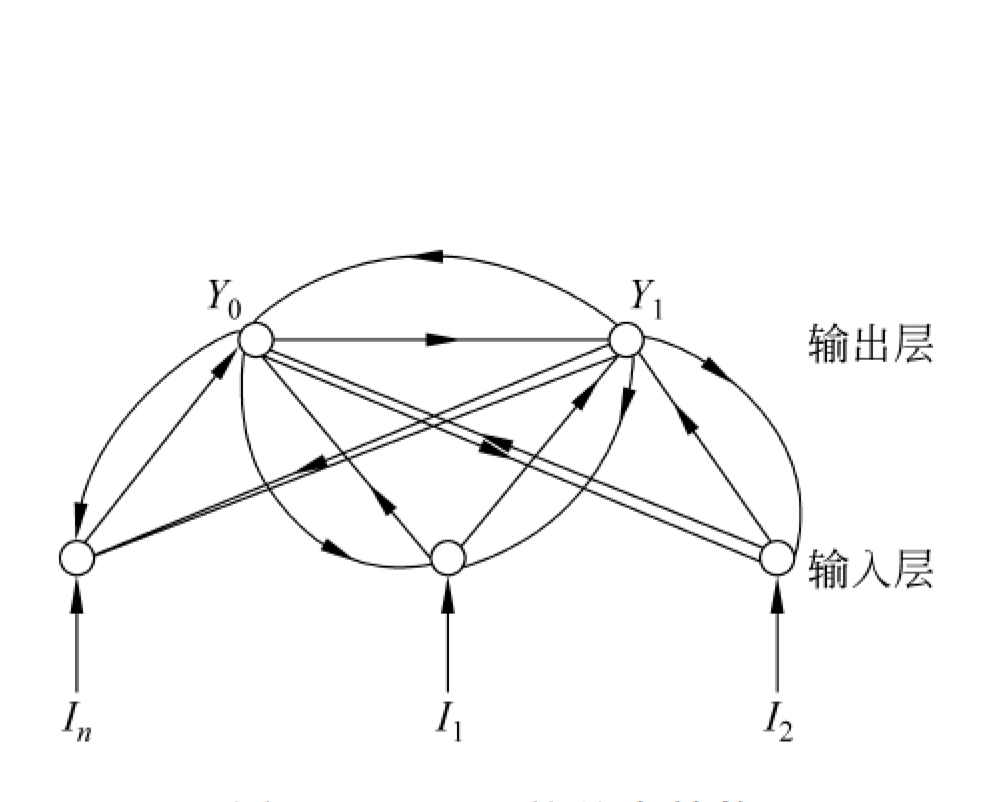



图3.9 ART的基本结构

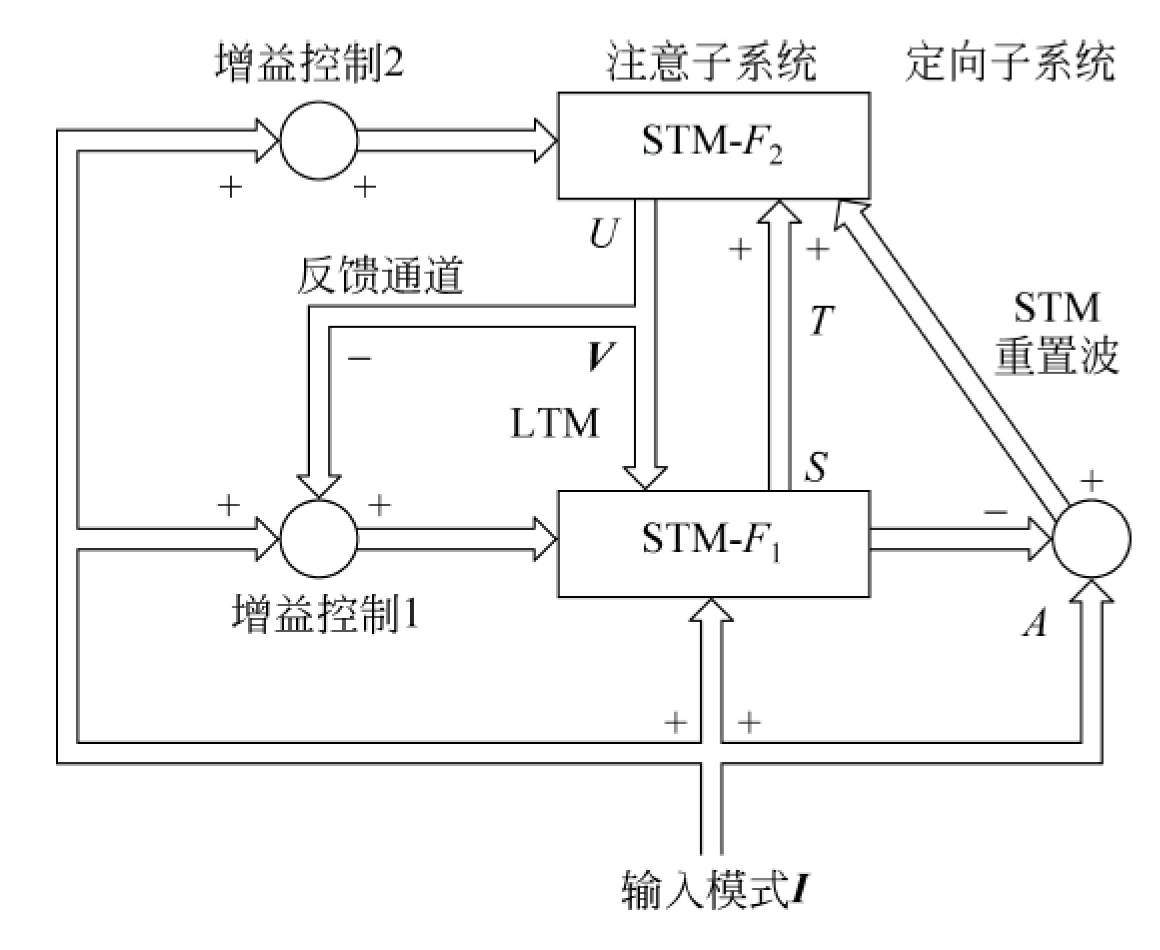



图3.10 ART模型的框图

它由两个子系统组成，一个称为注意子系统(Attentional Subsystem)，一个称为取向子系统(Orienting Subsystem)，也称调整子系统。这两个子系统是功能互补的子系统。ART模型就是通过这两个子系统和控制机制之间的交互作用来处理熟悉的事件或不熟悉的事件。在注意子系统中，有 $F_1$、$F_2$ 这两个用短时记忆单元组成的部件，即 STM-$F_1$ 和 STM-$F_2$。在 $F_1$ 和 $F_2$ 之间的连接通道是长时记忆 LTM。增益控制有两个作用：一个作用是在 $F_1$ 中用于区别自下而上和自上而下的信号；另一作用是当输入信号进入系统时，$F_2$ 能够对来自 $F_1$ 的信号起阈值作用。调整子系统是由 A 和 STM 重置波通道组成。

注意子系统的作用是对熟悉事件进行处理。在这个子系统中建立熟悉事件对应的内部表示，以便响应有关熟悉事件，这实际上是对 STM 中的激活模式进行编码。同时，在这个子系统中还产生一个从 $F_2$ 到 $F_1$ 的自上而下的期望样本，以帮助稳定已被学习了的熟悉事件的编码。

调整子系统的作用是对不熟悉事件产生响应。在有不熟悉事件输入时，孤立的一个注意子系统无法对不熟悉的事件建立新的聚类编码；故而设置一个调整子系统，当有不熟悉事件输入时，调整子系统马上产生重置波对 $F_2$ 进行调整，从而使注意子系统对不熟悉事件建立新的表达编码。实际上，当自下而上的输入模式和来自 $F_2$ 的自上而下的引发模式，即期望在 $F_1$ 中不匹配时，调整子系统就会发出一个重置波信号到 $F_2$，它重新选择 $F_2$ 的激活单元，同时取消 $F_2$ 原来所发出的输出模式。

简而言之，注意子系统的功能是完成由下向上的向量的竞争选择，以及完成由下向上向量和由上向下向量的相似度比较。而取向子系统的功能是检验期望向量模式 $\boldsymbol{V}$ 和输入模式 $\boldsymbol{I}$ 的相似程度；当相似度低于某一给定标准值时，即取消该时的竞争优胜者，转而从其余类别中选取优胜者。

ART 模型就是由注意子系统和调整子系统共同作用，完成自组织过程的。

### 3.5.2 ART 的基本工作原理

在 ART 模型中，其工作过程采用 2/3 规则的。所谓 2/3 规则，就是在 ART 网络中，三个输入信号中要有两个信号起作用才能使神经元产生输出信号。ART 网络的整个工作过程中，2/3 规则都在起作用。在说明 ART 模型的工作原理之前，先介绍 2/3 规则。

#### 1. 2/3 规则

考虑如图 3.10 所示的 ART 模型，很明显在 $F_1$ 层中，有三个输入信号源：输入信号 $\boldsymbol{I}$，增益控制输入，自上而下的模式输入。所谓 2/3 规则，就是指 $F_1$ 中这三个输入信号对 $F_1$ 的激发作用和关系。

2/3 规则可以用如图 3.11 所示的图形来说明。

(1) 自上而下的单输入情况

这时的情况如图 3.11(a)所示。$F_1$ 从三个输入信号源中，只接收来自 $F_2$ 的自上而下的引发模式。故而，$F_1$ 中的三个输入信号源中，并没有两个输入信号源起作用，而只有一个输入信号源即来自 $F_2$ 的自上而下的引发模式。所以 $F_1$ 中的神经元不会被激活，$F_1$ 的神经元不会产生信号输出。

(2) 自下而上的双输入情况

这时的情况如图3.11(b)所示。这时,在 $F_1$ 的三个输入信号源中,有输入信号 $\boldsymbol{I}$ 进行输入,并且有 $\boldsymbol{I}$ 通过增益控制后所产生的对 $F_1$ 的输入;由于这两个输入信号起作用,故而 $F_1$ 中的神经元被激活,$F_1$ 能产生信号输出。

(3) 自下而上输入及自上而下引发模式输入的情况

这时的情况如图3.11(c)所示。它说明了自下而上输入模式和自上而下的引发模式共同作用于 $F_1$ 的过程。这个过程也就是输入模式 $\boldsymbol{I}$ 和来自 $F_2$ 的自上而下的引发模式匹配过程。在这时,$F_2$ 的输出信号会加到增益控制中对其中的输入信号 $\boldsymbol{I}$ 产生抑制作用,所以,增益控制不会产生信号送去 $F_1$。在 $F_1$ 中,同时接收到自下而上输入信号以及自上而下的 $F_2$ 输出信号的神经元才会被激活;而只接收到其中一个信号的神经元则不会被激活。

(4) 模态竞争情况

当注意子系统从一个模态向另一个模态转移时,在这个转移的瞬间会禁止 $F_1$ 被激活,因为,这是一个过渡过程,它不反映模式的实质内容。故 $F_1$ 不能被激活。模态竞争的情况如图3.11(d)所示。

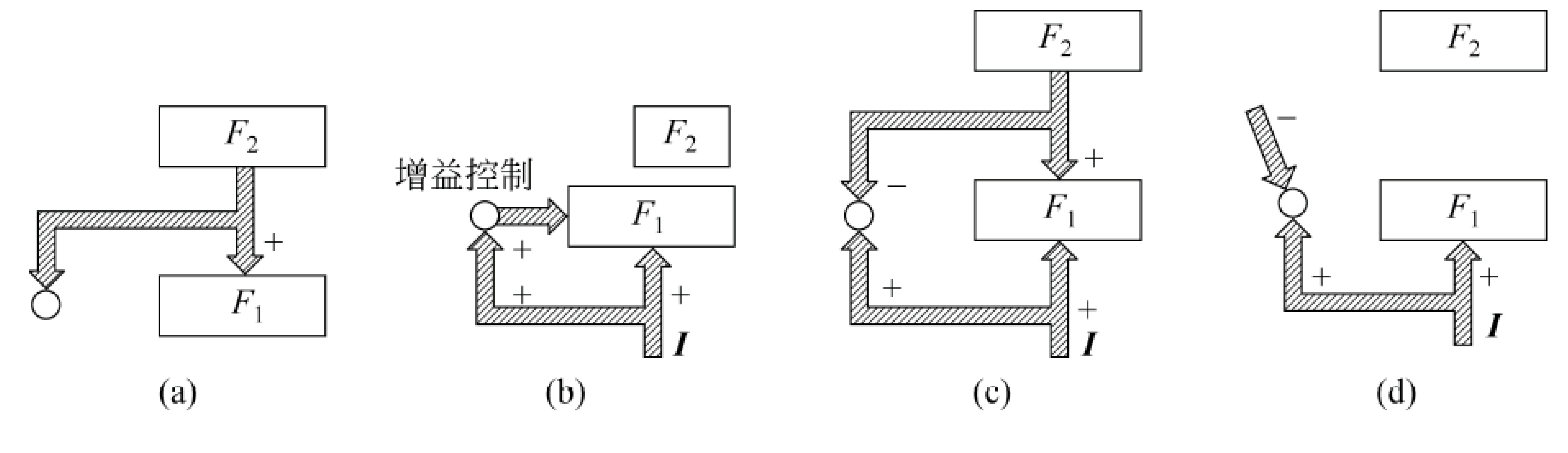



图3.11　2/3规则

## 2. ART模型的基本工作原理

在ART模型中,显然分为 $F_1$、$F_2$ 两层神经网络。对于注意子系统,$F_1$ 和 $F_2$ 这两层的作用可以用图3.12表示。

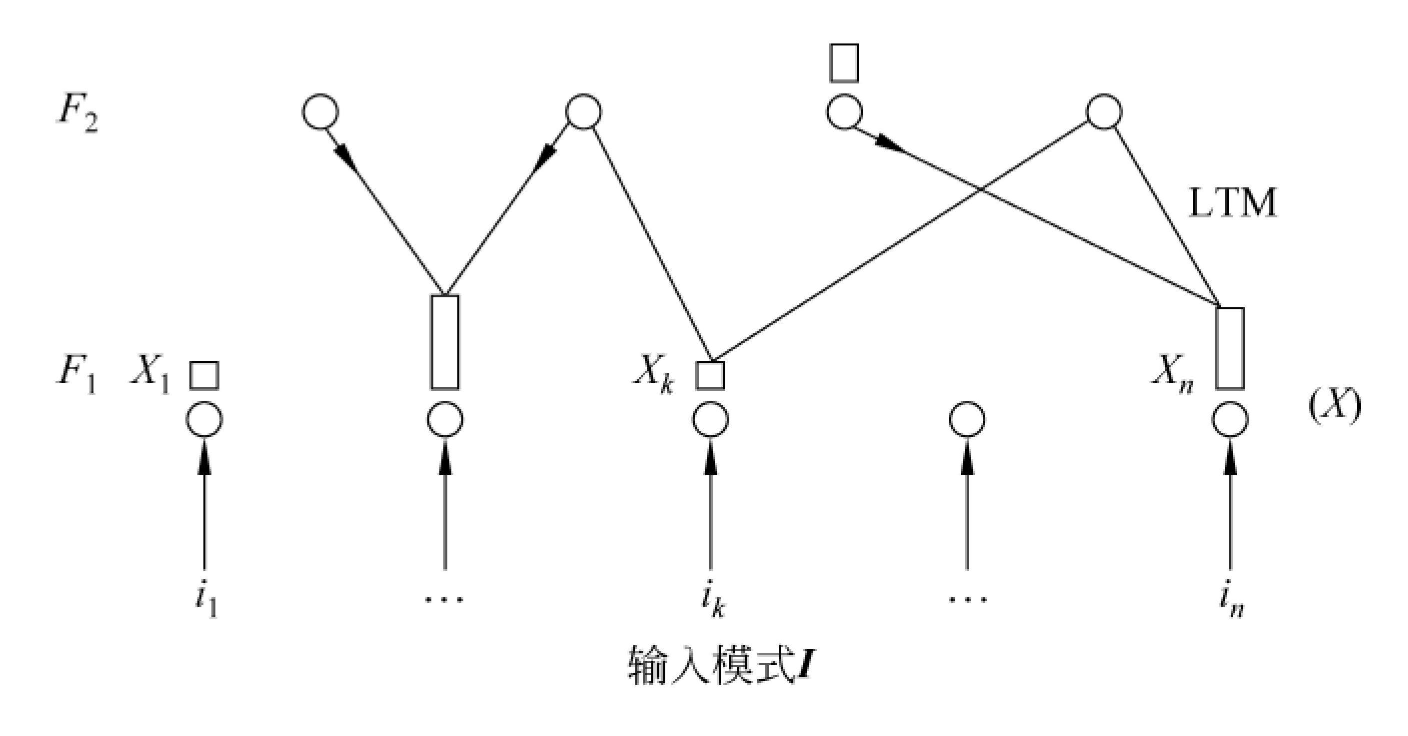



图3.12　$F_1$ 和 $F_2$ 层的信息处理

$F_1$ 层接收输入模式 $\boldsymbol{I}$,则在 $F_1$ 中被转换成激活模式 $\boldsymbol{X}$,$\boldsymbol{X}$ 由 $F_1$ 中的激活神经元表示,如图3.12中的长方形所示。这个模式 $\boldsymbol{X}$ 被短期存储在 $F_1$ 中。只有激活值足够高的神经元才能产生输出信号并通过连接传送到 $F_2$ 的神经元去。

在 $F_1$ 中，由 **X** 所产生的 $F_1$ 输出模式为 **S**，**S** 模式通过连接送到 $F_2$ 的神经元输入端，并在 $F_2$ 的神经元的输入端产生一个和 **S** 不同的模式 **T**。从 **S** 到 **T** 的转换称为自适应滤波。无论 $F_1$ 还是 $F_2$，其神经元是一般形式的神经元结构。一般而言，这些神经元的状态，输入和输出并不相同。

在 $F_2$ 层中，模式 **T** 经过 $F_2$ 神经元的相互作用会迅速地被转换。这个相互作用是对输入模式 **T** 的比较及除弱增强的过程。其结果产生一个短期存储在 $F_2$ 中的模式 **Y**，这也是 $F_2$ 的状态。

在一般情况时，从 **T** 到 **Y** 这个比较，除弱增强的转换会使多个神经元处于激活状态。这时，这种转换结果变为由 $F_2$ 中的多个神经元群来表达。这个转换过程自动地把 $F_1$ 的输入模式 **I** 划分到不相交的各个识别聚类中去，每个类对应于 $F_2$ 中的某个特征神经元。在特殊情况下，从 **T** 到 **Y** 的比较、除弱增强过程就是在 $F_2$ 中选择一个与当前输入 **I** 相对应的而输出值最大的神经元的过程，所选择的神经元就是用于表示激活模式的唯一神经。为了说明 ART 模型有关工作的基本原理，下面分 5 点进行介绍。

(1) 自下而上的自适应滤波和 STM 中的对比度增强过程。

输入信号 **I** 加到注意子系统的 $F_1$ 的输入端，经过 $F_1$ 的节点变换成激活模式 **X**，这一过程起到特征检出作用。在 $F_1$ 中，激活值较高的神经元就会有输出到 $F_2$ 的信号，并成为输出模式 **S**。**S** 经过 $F_1$ 到 $F_2$ 的连接通道时受到加权组合(LTM)，变换成模式 **T** 后作用于 $F_2$ 的输入端。**S** 到 **T** 的变换称为自适应滤波。$F_2$ 接收到 $T$ 后通过神经元间的相互作用迅速产生对比度增强了的激活模式 **Y**，并且存储于 $F_2$ 中。如图 3.13(a)所示。

这个阶段的学习是一个变换系列：**I**—**X**—**S**—**T**—**Y**。

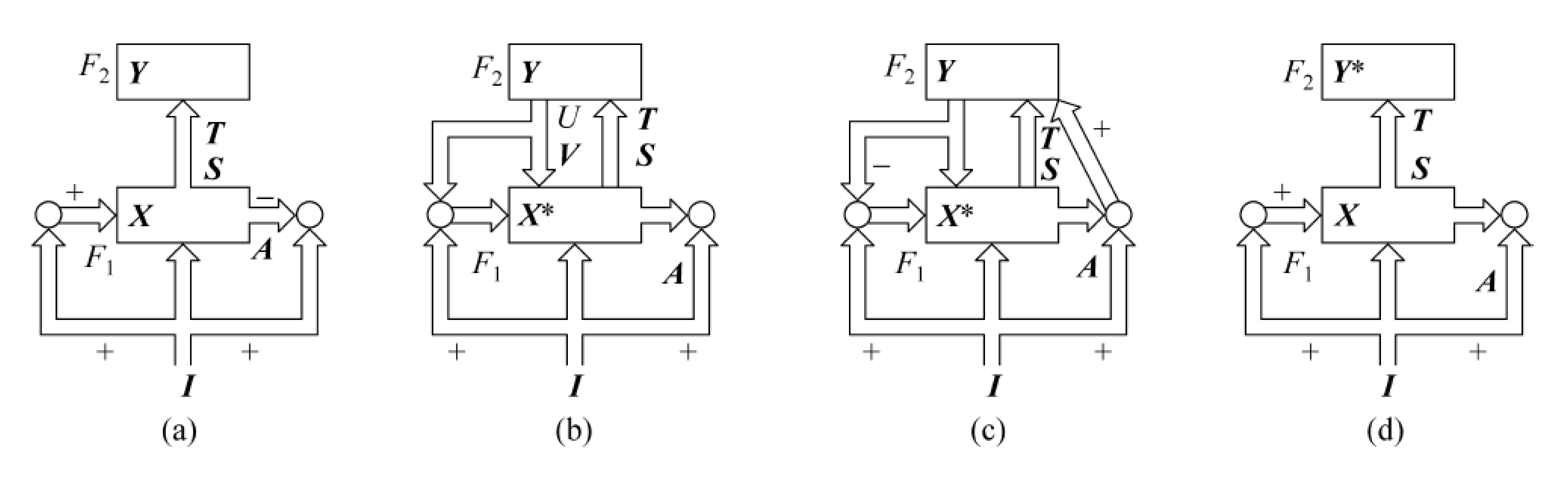



图 3.13 ART 模型的工作过程

(2) 自上而下的学习期望匹配和对已学习编码的稳定。

一旦当自下而上的变换 **X**—**Y** 完成之后，**Y** 就会产生自上而下的输出信号的模式 **U**，并送向 $F_1$，只有激活值足够大的才会向反馈通道送出信号 **U**。**U** 经加权组合变换成模式 **V**。**V** 称为自上而下的模板，或学习期望。

由于 **X** 和 **V** 这两个模式对 $F_1$ 输入，则它们的共同作用在 $F_1$ 中产生激活模式 $\boldsymbol{X}^*$。一般而言，$\boldsymbol{X}^*$ 和只由输入信号产生的 $X$ 模式不同。这时，$F_1$ 的作用就是试图使 **V** 和 **I** 匹配，其匹配结果确定了以后的作用过程。这时的情况如图 3.13(b)所示。

(3) 注意子系统和取向子系统相互作用过程。

这个过程和输入 **I** 有关。在图 3.13(a)中，在输入模式 **I** 产生 **X** 的同时，也会激活取向子系统 **A**；只是在 **A** 产生输出之前，$F_1$ 中的 **X** 对 **A** 所产生的输出端起禁止作用。当 $F_2$ 的

反馈模式 $\boldsymbol{V}$ 与 $F_1$ 的输入模式失配时,就会大大减弱这一禁止作用,当减弱到一定的程度时,$\boldsymbol{A}$ 就被激活。如图 3.13(c)所示。

$\boldsymbol{A}$ 被激活之后就向 $F_2$ 送出重置信号,并作用于 $F_2$ 的全部神经元,从而改变 $F_2$ 的状态,取消了原来的自上而下的学习期望 $\boldsymbol{V}$;终止了 $\boldsymbol{V}$ 和 $\boldsymbol{I}$ 的失配;于是输入 $\boldsymbol{I}$ 再次作用直到 $F_2$ 产生新的状态 $\boldsymbol{Y}^*$。如图 3.13(d)所示。

$\boldsymbol{Y}^*$ 会产生新的自上而下的学习期望 $\boldsymbol{V}^*$,如果 $\boldsymbol{V}^*$ 仍然和 $\boldsymbol{I}$ 失配,那么,取向子系统 $\boldsymbol{A}$ 继续起作用;这样,产生一个快速的一系列匹配与重置过程。这个过程控制 LTM 的搜索从而调整了 LW 对外界环境的编码。这个过程一直执行下去,直到 $F_2$ 送出的模式 $\boldsymbol{V}$ 和输入 $\boldsymbol{I}$ 相互匹配为止。

(4) 需考虑的一些特点。

在注意子系统的增益控制及起动这一自上而下的学习期望匹配过程中,还应考虑一些有关特点。

例如在 $F_1$ 输出自下而上的作用之前,$F_2$ 已被激活。这时 $F_2$ 就会产生自上而下的学习期望并作用于 $F_1$;这时 $F_1$ 也会被激活,并产生自下而上的作用过程。显然,需要对来自外部输入的激活以及来自 $F_2$ 的反馈激活进行区分。所以,设置一个辅助机构进行区分激活来源的工作。这个辅助机构称为注意增益控制。

为 $F_2$ 被激活时,注意起动机构会向 $F_1$ 发送学习期望信号,注意增益控制就会给出禁止作用,从而影响 $F_1$ 对输入响应灵敏度,使得 $F_1$ 可以区分激活信号的来源。

(5) 匹配。

采用 2/3 规则,以确定 $F_1$ 的输出。这实际上是把存储模式和输入模式进行匹配的规则。

### 3. ART 模型的工作过程

在图 3.10 所示的 ART 结构中,$F_1$ 可称为比较层,$F_2$ 可称为识别层。

比较层 $F_1$ 接收输入模式 $\boldsymbol{I}$,初始时不作任何变动作为输出向量 $\boldsymbol{S}$ 送去识别层 $F_2$;此后,$F_1$ 同时接受识别层输出的向量 $\boldsymbol{V}$ 和输入模式 $\boldsymbol{I}$,还有增益控制的输出,并按 2/3 规则产生输出。在初始时,增益控制的输出为 $\boldsymbol{I}$,而 $\boldsymbol{V}$ 设置为 0,故有 $\boldsymbol{S}$ 等于输入 $\boldsymbol{I}$。

识别层 $F_2$ 是用作输入向量分类器的。在识别层中,只有一个神经元和输入的向量 $\boldsymbol{S}$ 最优匹配,这个神经元就会被激活,而其他神经元则被抑制。根据神经元的结构原理,最优匹配规则如下:

$$\sum_i W_{ic}\boldsymbol{S}_i = \max \sum_i \boldsymbol{W}_{ij}\boldsymbol{S}_i \tag{3.28}$$

其中,$\boldsymbol{S}$ 是输入 $F_2$ 的向量;$\boldsymbol{S}=(S_1,S_2,\cdots,S_n)$;$\boldsymbol{W}_j$ 是识别层中第 $j$ 个神经元和比较层中神经元从 $F_1$—$F_2$ 的权系数向量 $\boldsymbol{W}_j=(W_{1j},W_{2j},\cdots)$;$\boldsymbol{W}_c$ 是识别层中最优匹配神经元 $C$ 从 $F_1$—$F_2$ 的权系数向量 $\boldsymbol{W}c=(W_{1c},W_{2c},\cdots)$。

**注意**:最优匹配神经元 $C$ 到比较层神经元有从 $F_2$—$F_1$ 的权系数向量 $\boldsymbol{W}c'$,$\boldsymbol{W}c'=(W_{c1},W_{c2},\cdots)$。很明显,$\boldsymbol{W}c$ 和 $\boldsymbol{W}c'$ 就组成了输入向量的类别样本,也是权系数的形态表示一类模式。

在识别层中,为了使一个神经元有最大输出值取得竞争的优胜,并抑制其他神经元。故而识别层有横向连接,每个神经元的输出和正的权系数相乘后作为本神经元的一个输入,而其他神经元的输出和负权系数相乘后再作为本神经元的输入。这种作用等于加强自身,抑

制其他，从而保证了只有一个神经元被激活。这种情况如图 3.14 所示。

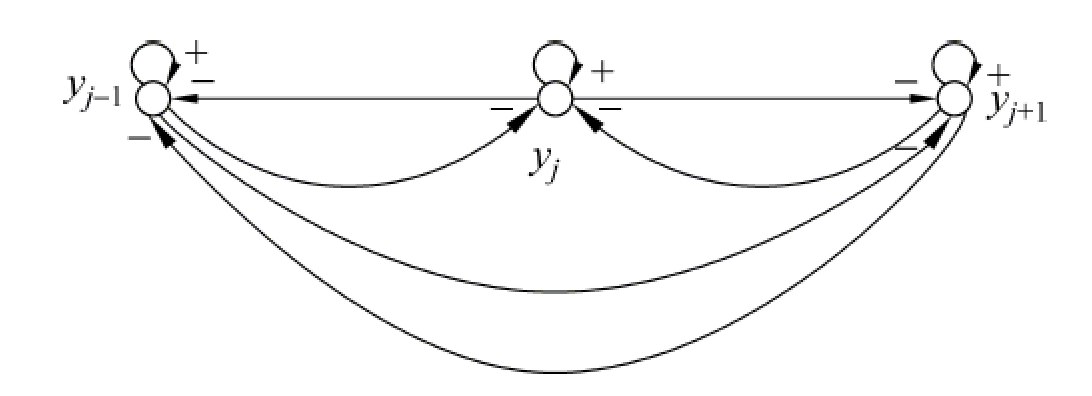



图 3.14　$F_2$ 层的横向连接

增益控制有两部分，它们的作用功能不同。识别层 $F_2$ 的增益控制输出原则为：只要输入向量 $\boldsymbol{I}$ 有一个元素为 1，则输出 1。比较层 $F_1$ 的增益控制原则为：只要在 $\boldsymbol{I}$ 有一个元素为 1，同时 $F_2$ 的输出向量 $\boldsymbol{U}$ 全部元素为 0 时，才输出 1。

重置作用是在输入信号 $\boldsymbol{I}$ 和 $F_1$ 的输出 $\boldsymbol{S}$ 之间的匹配存在问题，差别大于某警戒值时，则发清零信号到 $F_2$。以便重新进行识别。

ART 网络的学习分类过程分为三步，即识别、比较和搜索。下面作简要介绍。

(1) 识别

初始化时，网络无输入信号，故 $\boldsymbol{I}$ 全部元素为 0；识别层 $F_2$ 增益控制输出为 0，识别层 $F_2$ 输出全部为 0。在有模式 $\boldsymbol{I}$ 输入后，$\boldsymbol{I}$ 必有元素为 1，故 $F_1$ 增益控制、$F_2$ 增益控制均输出 1；比较层 $F_1$ 按 2/3 规则全部复制 $\boldsymbol{I}$ 作为输出；$\boldsymbol{S}=(S_1,S_2,\cdots,S_n)$。接着识别层 $F_2$ 的每个神经元 $j$ 执行下面操作；从而求出最优匹配神经元 $C$：

$$\sum_i \boldsymbol{W}_{ic}\boldsymbol{S}_i = \max \sum_i \boldsymbol{W}_{ij}\boldsymbol{S}_i$$

则神经元 $C$ 输出 1，其余输出 $\boldsymbol{U}$。这些输出送回比较层 $F_1$。$F_2$ 输出的值为 $\boldsymbol{U}=(U_1, U_2,\cdots)$。找寻最优匹配神经元 $C$ 的过程就是识别。

(2) 比较

从识别层 $F_2$ 反馈到 $F_1$ 的向量 $\boldsymbol{U}$ 不再全部为 0，$F_1$ 增益控制输出 0。按 2/3 规则，只有输入及反馈向量 $\boldsymbol{U}$ 的元素同时为 1，所激励的神经元才会被激活。从另一个角度讲，就是来自 $F_2$ 的反馈强迫输入向量 $\boldsymbol{I}$ 中那些不匹配存储模式 $\boldsymbol{U}$ 的 $\boldsymbol{S}$ 元素为 0。

如果 $\boldsymbol{I}$ 与 $\boldsymbol{U}$ 不匹配，则产生的 $\boldsymbol{S}$ 只有少数元素为 1，这也说明模式 $\boldsymbol{U}$ 不是所要寻找的 $\boldsymbol{I}$ 模式。取向子系统对 $\boldsymbol{I}$ 和 $\boldsymbol{S}$ 的匹配程度进行判别，如果低于给定的警戒值，则发出重置信号，使识别层 $F_2$ 激活的神经元清零；这也说明该神经元失去竞争的资格。到此这个分类阶段比较过程结束。如果 $\boldsymbol{I}$ 与 $\boldsymbol{U}$ 匹配，则输入模式 $\boldsymbol{I}$ 所属的类别已找到，分类结束。

(3) 搜索

在 $\boldsymbol{I}$ 与 $\boldsymbol{U}$ 不匹配时，为了找到较好的匹配必须对其余的模式进行搜索。重置信号把识别层 $F_2$ 的神经元全部清 0，则 $F_1$ 增益控制又输出 1，网络返回到初始状态。输入模式 $\boldsymbol{I}$ 再进行输入，识别层的另一个神经元会取得优胜，则反馈一个新的存储模式 $\boldsymbol{U}$ 送回比较层 $F_1$。接着又进行匹配比较，如不匹配，则又重置识别层……不断执行下去。搜索过程直到产生下列情况之一才会停止。

① 找到一个存储模式，在警戒值范围内和输入模式 $\boldsymbol{I}$ 匹配，则 ART 网络进入学习阶段。修正和匹配神经元 $C$ 相关的权系数 $\boldsymbol{W}_{ic}$ 和 $\boldsymbol{W}_{ci}$。

② 搜索了全部模式后，没有一个模式能够和 $\boldsymbol{I}$ 相似匹配，则网络也进入学习阶段。把原

来分配模式的神经元 $j$ 赋于输入式 $\boldsymbol{I}$,构造相应的权系数 $\boldsymbol{W}_{ij}$ 和 $\boldsymbol{W}_{ji}$,并作为样本模式存储。

特别应指出的是:搜索过程包含了识别和比较二个阶段。搜索是识别-比较-识别-比较的多次重复。

严格地说,ART 应分成搜索和学习这两种最主要的过程和功能。

### 3.5.3 ART 模型的数学描述

在 ART 模型中,$F_1$ 或 $F_2$ 中的神经元用 $N_k$ 表示,神经元被激活后产生的激活值用 $X_k$ 表示,从神经生理学的研究结果,可以知道神经元的激活值,即神经元未经 S 函数处理的输出 $X_k$ 满足下面的微分方程:

$$\varepsilon\frac{\mathrm{d}X_k}{\mathrm{d}t}=-X_k+(1-AX_k)J_k^{+}+(B+CX_k)J_k^{-} \tag{3.29}$$

其中,$\varepsilon$ 是远小于 1 的正实数,$J_k^{+}$ 是送到神经元 $N_k$ 的所有激励输入之和,$J_k^{-}$ 是送到神经元 $N_h$ 的所有抑制输入之和,$A,B,C$ 是非负常数,$X_k$ 的取值范围为$[-BC-1,A-1]$。

#### 1. $F_1$ 层的数学描述

用 $N_i$ 表示 $F_1$ 的神经元,并且 $i=1,2,\cdots,n$,则有

$$\varepsilon\frac{\mathrm{d}X_i}{\mathrm{d}t}=-X_i+(1-A_1X_i)J_i^{+}+(B_1+C_1X_i)J_i^{-} \tag{3.30}$$

很明显,有 $F_1$ 的激活模式:

$$X=\{X_1,X_2,\cdots,X_n\}$$

(1) $J_i^{+}$ 的形式

由于 $F_1$ 神经元 $N_i$ 的激励输入 $J_i^{+}$ 是自下而上的输入 $\boldsymbol{I}_i$ 以及自上而下的输入 $\boldsymbol{V}_i$ 之和,故有 $J_i^{+}=\boldsymbol{I}_i+\boldsymbol{V}_i$,其中 $\boldsymbol{I}_i$ 是一个 $n$ 维输入向量,$\boldsymbol{I}=\{I_1,I_2,\cdots,I_n\}$;

$$\boldsymbol{V}_i=D_1\sum_i f(X_i)W_{ji} \tag{3.31}$$

这里,$f(X_j)$是 $F_2$ 中神经元 $N_i$ 的输出,$\boldsymbol{W}_{ji}$是 $N_j$ 到 $N_i$ 的连接权系数;$D_1$ 是系数。$V=\{V_1,V_2,\cdots,V_n\}$,也是 $n$ 维向量。

(2) $J_i^{-}$ 的形式

对 $F_1$ 层,抑制输入 $J_i^{-}$ 是由注意子系统的增益控制信号来控制,即

$$J_i^{-}=\sum_i f(X_i)\quad j=n+1,n+2,\cdots,n+m \tag{3.32}$$

当且仅当 $F_2$ 的激活值很高时,$J_i^{-}=0$,否则 $J_i^{-}>0$。

#### 2. $F_2$ 层的数学描述

用 $N_j$ 表示 $F_2$ 的神经元,并且 $j=n+1,n+2,\cdots,n+m$,则有

$$\varepsilon\frac{\mathrm{d}X_j}{\mathrm{d}t}=-X_j+(1-A_2X_j)J_j^{+}+(B_2+C_2X_j)J_j^{-} \tag{3.33}$$

则有 $F_2$ 的激活模式 $Y=\{X_{n+1},X_{n+2},\cdots,X_{n+m}\}$。

选择 $F_2$ 中的激活模式的输入和参数,使到 $F_2$ 中具有来自 $F_1$ 的最大输入的神经元取得竞争的胜利。故而对 $J_j^{+}$ 和 $J_i^{-}$ 考虑应有如下形式。

(1) $J_j^+$ 的形式

$$J_j^+ = g(X_j) + T_j \tag{3.34}$$

其中,$g(X_j)$为 $N_j$ 的自反馈信号,$T_j$ 是从 $F_1$ 来的到 $F_2$ 的输入模式:

$$T_j = D_2 \sum_i h(X_i) W_{ji} \tag{3.35}$$

这里的 $h(X_i)$是 $F_1$ 中神经元 $N_i$ 的输出,$D_2$ 是系数,$W_{ji}$是 $F_1$ 到 $F_2$ 的神经元的连接权系数。

(2) $J_i^-$ 的形式

$$J_j^- = \sum_{k \neq j} g(X_k) \tag{3.36}$$

迄今为止,按生物神经网络(Biological Neural Networks,BNN)巨量并行分布方式构造的各种人工神经网络,虽然已经在信息处理中扮演着越来越重要的角色,但是并没有显示出人们所期望的聪明智慧来。对以仿效大脑神经系统为目的的人工神经网络的研究历程进行一些分析和反思,探讨下一步可能采取的方法步骤,对今后智能信息科学的进一步发展将是有益的。要让 ANN 更好更快地向 BNN 学习,就有必要对今后 ANN 的主攻方向、研究路线、方法步骤、关键技术和应当采取的措施等方面,作一些考虑和调整。

概括起来,为了使 ANN 向 BNN 学习得更好更快,可以从以下几个方面着手。

(1) 明确 ANN 的主要智能优势是擅长非精确性信息处理。按照输入和输出特性的不同,可将智能系统最经常处理的信息类型归纳为四类。其中,数值计算和逻辑推理所对应的信息处理任务,现行的冯·诺依曼数字计算机有着成熟而巨大的能力,ANN 既无必要也无优势在此等领域中与之竞争。其他三类都在输入信息或输出信息中包含有非精确信息处理的内容,因而它们是 ANN 能发挥作用的领域。人类的绝大部分脑力劳动正是投入在这三类信息处理上。

(2) 探索新的 ANN 体系结构。神经解剖学方面的研究表明,在 BNN 中除了有由神经细胞体-轴突-突触-树突构成的神经电位脉冲电路系统外,还有一个在前后突触间约 0.02$\mu$s 的间隙中释放化学性神经递质(谷氨酸)的调节系统,这个化学递质系统的作用相当于半导体的栅极电路,起着至关紧要的调控放大作用。因此,如何建立神经电脉冲系统和化学递质系统合成的耦合系统数学模型,分析其工作机理并提出可行简便算法,是从建模上使 ANN 更靠近 BNN 的一条值得重视的途径。

(3) 寻求新的网络拓扑结构和相应的学习算法。迄今为止,以和-积式神经元为基础的前馈型多层(特别是三层)神经网络,与误差反向传播学习算法相配合的 ANN 结构体制,获得了最广泛的应用。但是,它在性能上仍存在若干待改进的地方。例如,它只能调整权值,不能调整网络拓扑结构,无法实现注意力集中功能;学习新样本时,会“冲乱”原已学好保存下来的旧样本;其学习算法中,包含有较复杂的非线性激活函数的求导运算过程等。

## 3.6 神经网络集成

1990 年汉森(L K Hansen)和萨拉蒙(P Salamon)提出了神经网络集成(Neural Network Ensemble)方法[305]。他们证明,可以简单地通过训练多个神经网络并将其结果进

行拟合,显著地提高神经网络系统的泛化能力。神经网络集成可以定义为用有限个神经网络对同一个问题进行学习,集成在某输入示例下的输出由构成集成的各神经网络在该示例下的输出共同决定。对神经网络集成的理论分析与其实现方法分为两个方面,即对结论生成方法以及对网络个体生成方法。

### 3.6.1 结论生成方法

汉森和萨拉蒙证明,对神经网络分类器来说,采用集成方法能够有效提高系统的泛化能力[305]。假设集成由 $N$ 个独立的神经网络分类器构成,采用绝对多数投票法,再假设每个网络以 $1-p$ 的概率给出正确的分类结果,并且网络之间错误不相关,则该神经网络集成发生错误的概率 $p_{err}$ 为

$$p_{err} = \sum_{k>N/2}^{N} \binom{N}{k} p^k (1-p)^{N-k} \tag{3.37}$$

在 $p<1/2$ 时,$p_{err}$ 随 $N$ 的增大而单调递减。因此,如果每个神经网络的预测精度都高于 50%,并且各网络之间错误不相关,则神经网络集成中的网络数目越多,集成的精度就越高,当 $N$ 趋向于无穷时,集成的错误率趋向于 0。在采用相对多数投票法时,神经网络集成的错误率比式(3.37)复杂得多,但是汉森和萨拉蒙的分析表明,采用相对多数投票法在多数情况下能够得到比绝对多数投票法更好的结果。

1995 年,克罗夫(A Krogh)和弗德尔斯毕(J Vedelsby)给出了神经网络集成泛化误差计算公式[389]。假设学习任务是利用 $N$ 个神经网络组成的集成对 $f: \Re^n \to \Re$ 进行近似,集成采用加权平均,各网络分别被赋以权值 $w_\alpha$,并满足式(3.38)和式(3.39)。

$$w_\alpha > 0 \tag{3.38}$$

$$\sum_\alpha w_\alpha = 1 \tag{3.39}$$

再假设训练集按分布 $p(x)$ 随机抽取,网络 $\alpha$ 对输入 $X$ 的输出为 $V^\alpha(X)$,则神经网络集成的输出为

$$\bar{V}(X) = \sum_\alpha w_\alpha V^\alpha(X) \tag{3.40}$$

神经网络 $\alpha$ 的泛化误差 $E^\alpha$ 和神经网络集成的泛化误差 $E$ 分别为

$$E^\alpha = \int \mathrm{d}x p(x)(f(x) - V^\alpha(x))^2 \tag{3.41}$$

$$E = \int \mathrm{d}x p(x)(f(x) - \bar{V}(x))^2 \tag{3.42}$$

各网络泛化误差的加权平均为

$$\bar{E} = \sum_\alpha w_\alpha E^\alpha \tag{3.43}$$

神经网络 $\alpha$ 的差异度 $A^\alpha$ 和神经网络集成的差异度 $\bar{A}$ 分别为

$$A^\alpha = \int \mathrm{d}x p(x)(V(x) - \bar{V}(x)^2 \tag{3.44}$$

$$\bar{A} = \sum_\alpha w_\alpha A^\alpha \tag{3.45}$$

则神经网络集成的泛化误差为

$$E = \bar{E} - \bar{A} \tag{3.46}$$

式(3.46)中的 $\bar{A}$ 度量了神经网络集成中各网络的相关程度。若集成是高度偏置的，即对于相同的输入，集成中所有网络都给出相同或相近的输出，此时集成的差异度接近于 0，其泛化误差接近于各网络泛化误差的加权平均。反之，若集成中各网络是相互独立的，则集成的差异度较大，其泛化误差将远小于各网络泛化误差的加权平均。因此，要增强神经网络集成的泛化能力，就应该尽可能地使集成中各网络的误差互不相关。

### 3.6.2　个体生成方法

1997 年，弗洛德(Y Freund)和沙皮尔(R E Schapire)以 AdaBoost 为代表，对 Boosting 类方法进行了分析[244]，并证明此类方法产生的最终预测函数 H 的训练误差满足式(3.47)，其中 $\varepsilon_t$ 为预测函数 $h_t$ 的训练误差，$\gamma_t = 1/2 - \varepsilon_t$。

$$H = \prod_t \left[2\sqrt{\varepsilon_t(1-\varepsilon_t)}\right]$$

$$= \prod_t \sqrt{1-4\gamma_t^2} \leqslant \exp\left(-2t\sum_t \gamma_t^2\right) \tag{3.47}$$

从式(3.47)可以看出，只要学习算法略好于随机猜测，训练误差将随 $t$ 以指数级下降。

1996 年，布雷曼(L Breiman)对 Bagging 进行了理论分析[101]。他指出，分类问题可达到的最高正确率以及利用 Bagging 可达到的正确率分别如式(3.48)和式(3.49)所示，其中 $C$ 表示序正确的输入集，$C'$ 为 $C$ 的补集，$I(\cdot)$ 为指示函数(Indicator Function)。

$$r^* = \int \max_j P(j \mid x) P_X(x) \tag{3.48}$$

$$r_A = \int_{x\in C} \max_j P(j \mid x) P_x(\mathrm{d}x) + \int_{x\in C'} \left[\sum_j I(\phi_A(x) = j) P(j \mid x)\right] P_X(x) \tag{3.49}$$

显然，Bagging 可使序正确集的分类正确率达到最优，单独的预测函数则无法做到这一点。

## 3.7　脉冲耦合神经网络

近年来，随着生物神经学的研究和发展，艾克霍恩(R Eckhorn)等通过对小型哺乳动物大脑视觉皮层神经系统工作机理的仔细研究，提出了一种崭新的网络模型——脉冲耦合神经网络模型(Pulse-Coupled Neural Network，PCNN)。PCNN 来源于对哺乳动物猫的视觉皮层神经细胞的研究成果，具有同步脉冲激发现象、阈值衰减及参数可控性等特性。由于其具有生物学特性的背景、以空间邻近和亮度相似集群的特点，因此在数字图像处理等领域具有广阔的应用前景。将 PCNN 的最新理论研究成果与其他新技术相结合，开发出具有实际应用价值的新算法是当今神经网络研究的主要方向之一。

1952 年，霍奇金与哈斯利(A F Huxley)开始研究神经元电化学特性[324]。1987 年，格雷(C M Gray)等发现猫的初生视觉皮层有神经激发相关振荡现象[325,324]。1989 年，艾克霍恩和格雷研究了猫的视觉皮层，提出了具有脉冲同步发放特性的网络模型[201,283]。1990 年，艾克霍恩根据猫的大脑皮层同步脉冲发放现象，提出了展示脉冲发放现象的连接模型[202]。对猴的大脑皮层进行的试验中，也得到了相类似的试验结果。1994 年，约翰逊(J L

Johnson)发表论文,阐述了 PCNN 的周期波动现象及在图像处理中具有旋转、可伸缩、扭曲、强度不变性[346]。通过对艾克霍恩提出的模型进行改进,就形成脉冲耦合神经网络(PCNN)模型。于 1999 年 IEEE 神经网络会刊出版了脉冲耦合神经网络专辑。国内也于 20 世纪 90 年代末开始研究脉冲耦合神经网络。

与传统方法相比,源自哺乳动物视觉皮层神经元信息传导模型的脉冲耦合神经网络是一种功能强大的图像处理工具,在解决图像处理具体应用时取得令人满意的性能。

## 3.7.1 Eckhorn 模型

1990 年,根据猫的视皮层的同步振荡现象,艾克霍恩提出一个脉冲神经网络模型[202],如图 3.15 所示。这个模型由许多相互连接的神经元构成,每个神经元包括两个功能上截然不同的输入部分:分别是常规的馈接(Feeding)输入,和起调制作用的连接(Linking)输入。而这两部分的关系并非像传统神经元那样是加耦合的关系,而是乘耦合的关系。

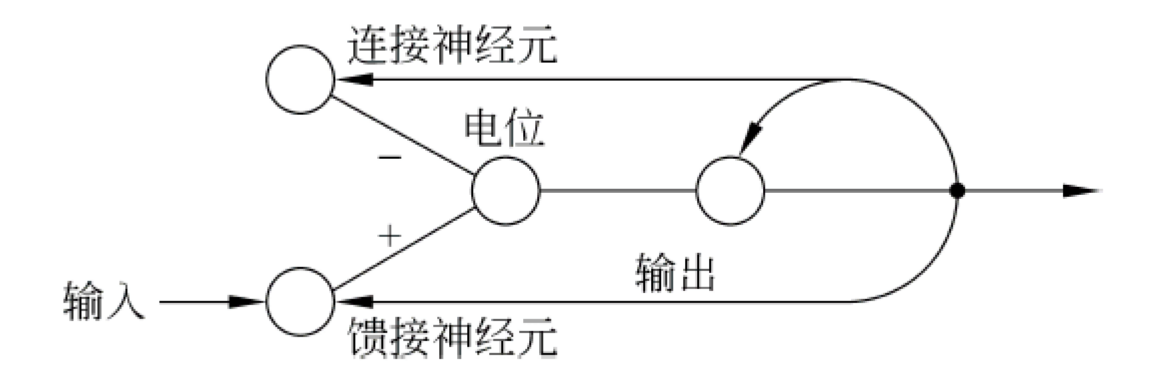



图 3.15 Eckhorn 神经元模型示意图

Eckhorn 模型可用如下方程描述:

$$U_{m,k} = F_k(t)[1 + L_k(t)] \tag{3.50}$$

$$F_k(t) = \sum_{i=1}^{N}[w_{ki}^f Y_i(t) + S_k(t) + N_k(t)] \otimes I(V^a, \tau^a, t) \tag{3.51}$$

$$L_k(t) = \sum_{i=1}^{N}[w_{ki}^l Y_i(t) + N_k(t)] \otimes I(V^l, \tau^l, t) \tag{3.52}$$

$$Y_k(t) = \begin{cases} 1, & U_{m,k}(t) \geqslant \theta_k(t) \\ 0, & \text{其他} \end{cases} \tag{3.53}$$

这里,一般表示为

$$X(t) = Z(t) \otimes I(v, \tau, t) \tag{3.54}$$

即

$$X[n] = X[n-1]\mathrm{e}^{-t/\tau} + VZ[n], \quad n = 1,2,\cdots,N \tag{3.55}$$

其中,$N$ 为神经元的个数,$w$ 为突触加权系数。当外部激励为 $S$ 型时,$Y$ 为二值输出。

## 3.7.2 脉冲耦合神经网络模型

由于 Eckhorn 模型提供了一个简单有效的方法来研究脉冲神经网络中的动态同步振荡活动,Eckhorn 模型的最大创新在于它引入了第二个感受野(Secondary Receptive Field),即连接域(Linking Field)。如果去掉连接输入部分,Eckhorn 模型中的神经元模型与常规的神经元模型没什么不同,而正是连接输入的引入,使我们对神经元如何整合输入有

了更深入的认识。通过对模型中神经元的电路进行分析，研究人员证明了：神经元的不同输入之间的关系不仅有加耦合的关系，而且有乘耦合的关系[341]。它很快被应用到图像处理领域，而它和它的许多变种模型被一起称为脉冲耦合神经网络(PCNN)。

图 3.16 给出了脉冲耦合神经元示意图。神经元主要有两个功能单元构成：馈接输入域和连接输入域，分别通过突触连接权值 $M$ 和 $K$ 来与其邻近的神经元相连。两功能单元都要进行迭代运算，迭代过程中按指数规律衰减。馈接输入域多加一个外部激励 $S$。可以用如下数学公式描述两个功能单元：

$$F_{ij}[n]=\mathrm{e}^{\alpha F\delta_n}F_{ij}[n-1]+S_{ij}+V_F\sum_{kl}\boldsymbol{M}_{ijkl}Y_{kl}[n-1] \tag{3.56}$$

$$L_{ij}[n]=\mathrm{e}^{\alpha L\delta_n}L_{ij}[n-1]+V_L\sum_{kl}\boldsymbol{K}_{ijkl}Y_{kl}[n-1] \tag{3.57}$$

式中，$F_{ij}$ 是第$(i,j)$个神经元的馈接，$L_{ij}$ 是耦合连接，$Y_{kl}$ 是$(n-1)$次迭代时神经元的输出。两个功能单元都要进行迭代运算，迭代过程按指数规律衰减。$V_F$ 和 $V_L$ 分别为 $F_{ij}$、$L_{ij}$ 的固有电位。这里 $\boldsymbol{M}$ 和 $\boldsymbol{K}$ 为连接权值系数矩阵，表示中心神经元受周围神经元影响的大小，反映邻近神经元对中心神经元传递信息的强弱，$M$ 和 $K$ 有多种取值选择方式，但选择要合适，一般不宜过大。

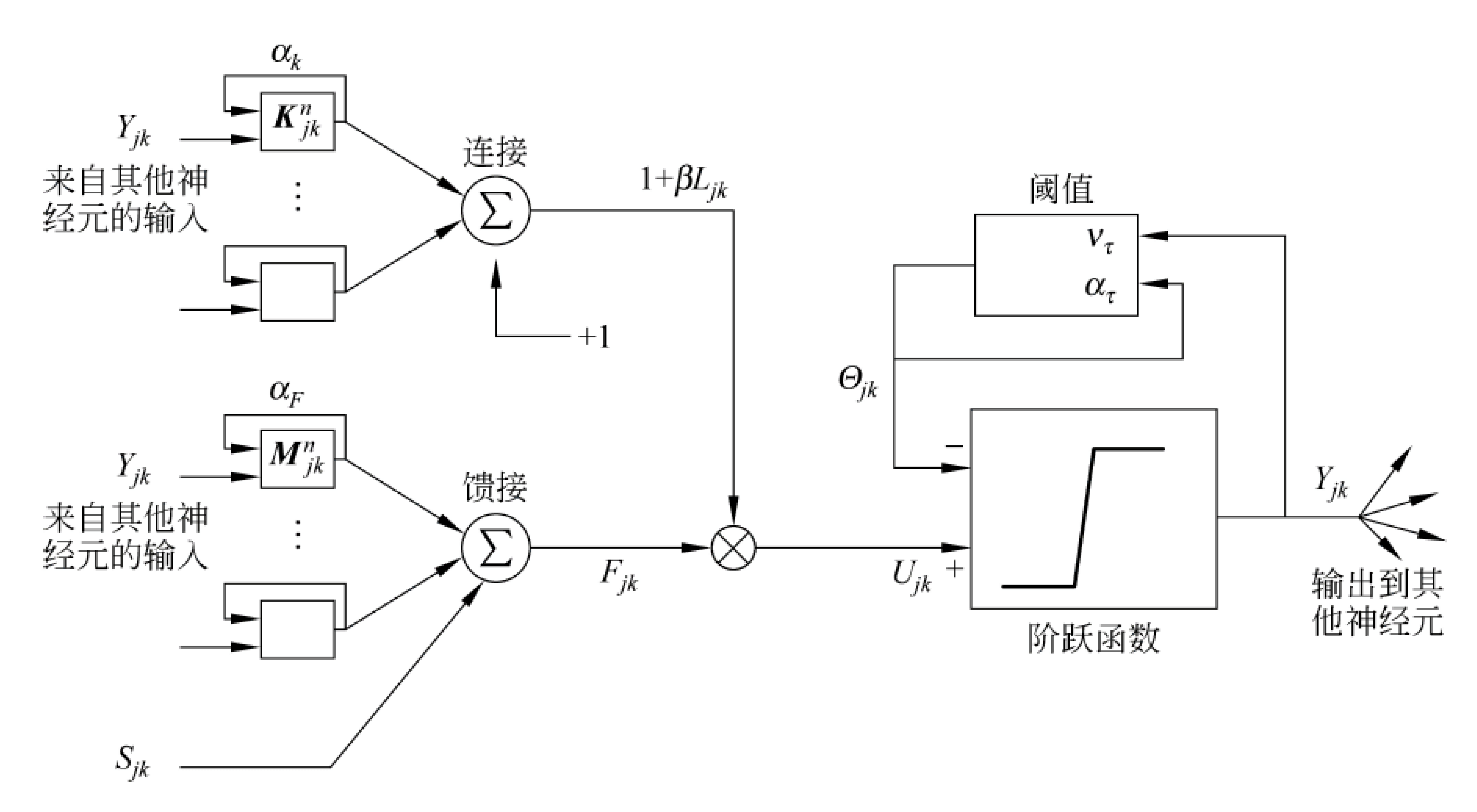



图 3.16　脉冲耦合神经元示意图

神经元内部活动项由这两个功能单元按非线性相乘方式共同组成，$\beta$ 为突触之间的连接强度系数。神经元内部活动项的数学表达式如下：

$$U_{ij}[n]=F_{ij}[n]\{1+\beta L_{ij}[n]\} \tag{3.58}$$

当神经元内部活动项大于动态阈值 $\Theta$ 时，产生输出时序脉冲序列 $Y$，即下式所示：

$$Y_{ij}[n]=\begin{cases}1, & U_{ij}[n]>\Theta_{ij}[n]\\ 0, & \text{其他}\end{cases} \tag{3.59}$$

动态阈值在迭代过程中衰减，当神经元激发兴奋$(U>\Theta)$时，动态阈值立刻增大，然后又按指数规律逐渐衰减，直到神经元再次激发兴奋。这个过程可描述为

$$\Theta_{ij}[n]=\mathrm{e}^{\alpha\Theta\delta n}\Theta_{ij}[n-1]+V_\Theta Y_{ij}[n] \tag{3.60}$$

式中 $\Theta$ 一般取一个比较大的值，相比 $U$ 的均值还大一个数量级。

PCNN由这些神经元排列(通常是矩阵)而成。$M$和$K$在神经元间传递信息通常是局部的,并符合高斯正态分布,但不必严格要求这样。矩阵$\boldsymbol{F}$、$\boldsymbol{L}$、$\boldsymbol{U}$、$\boldsymbol{Y}$初始化时,设其所有矩阵元素为零。$\Theta$元素的初始值可以是0,也可以根据实际需要设为某些更大值。任何有激励的神经元都将在第一次循环中激发兴奋,结果将生成一个很大的阈值。接下来需要经过几次循环才能使阈值衰减到足以使神经元再次激发兴奋。后者的情况趋向于围绕这些信息量小的初始循环。

本算法循环计算式(3.56)～式(3.60),直到用户决定停止。目前PCNN本身还没有自动停止的机制。

与传统神经网络相比,PCNN具有自己鲜明的特色,它具有如下特性。

(1) 变阈值特性:PCNN中各神经元之所以能动态发放脉冲,是因为它内部的变阈值函数作用的结果。由式(3.60)可见,它是随时间按指数规律衰减的。当神经元的内部行为$U$大于当前的阈值输出值时就发放。对于无连接耦合的PCNN来说,每一时刻的发放图就是对应于该阈值下的二值图像帧。对于存在连接耦合的PCNN来说,每一时刻的发放图就是对应于该阈值下带有捕获功能的二值图像帧。

(2) 捕获特性:PCNN的捕获过程就是使亮度强度相似输入的神经元能够同步发放脉冲,而同步的结果就好像把低亮度强度提升至先发放的那个神经元对应输入的亮度强度。这就意味着因捕获可使得某一神经元的先发放,而激励或带动邻近其他神经元提前点火。PCNN神经元间存在连接但不一定存在影响,存在影响但不一定存在连接,这一现象更加突出PCNN对突发事件的处理能力,表现在由于某种原因(如噪声)使得网络原本已经组织起的有序状态,因某个或某些神经元点火状态的改变而被打破时,网络可自动地适应新的变化,实现对信息的重新组织,达到一个新的有序状态。

(3) 动态脉冲发放特性:PCNN动态神经元的变阈值特性是其动态脉冲发放的根源,如果将由输入信号与突触通道的卷积和所产生的信号称为该神经元的(内部)作用信号,则当作用信号超过阈值时,该神经元被激活而产生高电平,又由于阈值受神经元输出控制,因此该神经元输出的高电平又反过来控制阈值的提高,从而作用信号在阈值以下,神经元又恢复为原来的抑制状态(即低电平)。这一过程在神经元输出上明显地形成一个脉冲发放。

(4) 同步脉冲发放特性:PCNN每个神经元有一个输入,并与其他神经元的输出有连接。当一个神经元发放时,它会将其信号的一部分送至与其相邻的神经元上。从而这一连接会引起邻近神经元比原来更快地点火,这样就导致了在图像的一个大的区域上产生同步振荡:以相似性集群产生同步脉冲发放,这一性质对于图像平滑、分割、图像自动目标识别、融合等具有重要的应用意义。

(5) PCNN时间序列:在点火捕获及脉冲传播特性的基础上,PCNN能够由二进制图像生成一维向量信息:$G[n]=\sum Y_{ij}[n]$。对时间序列信号进行分析,可以达到识别图像的目的。

### 3.7.3 贝叶斯连接域神经网络模型

与Eckhorn模型类似,我们提出的贝叶斯连接域神经网络(Bayesian Linking Field Network,BLFN)模型也是一个由众多神经元构成的网络模型,而且模型中的神经元都包含两类输入:一类是馈接(Feeding)输入,另一类是连接(Linking)输入,两类输入之间的耦合

关系是相乘。与 Eckhorn 模型不同的是：为了解决特征捆绑的问题，我们还引入了噪声神经元模型的思想、贝叶斯方法和竞争机制。

图 3.17 给出了我们模型中的一个神经元输入耦合方式的示意。由于模型中神经元的输出是发放概率，所以输入的耦合实际上是各个传入神经元的发放概率的耦合。

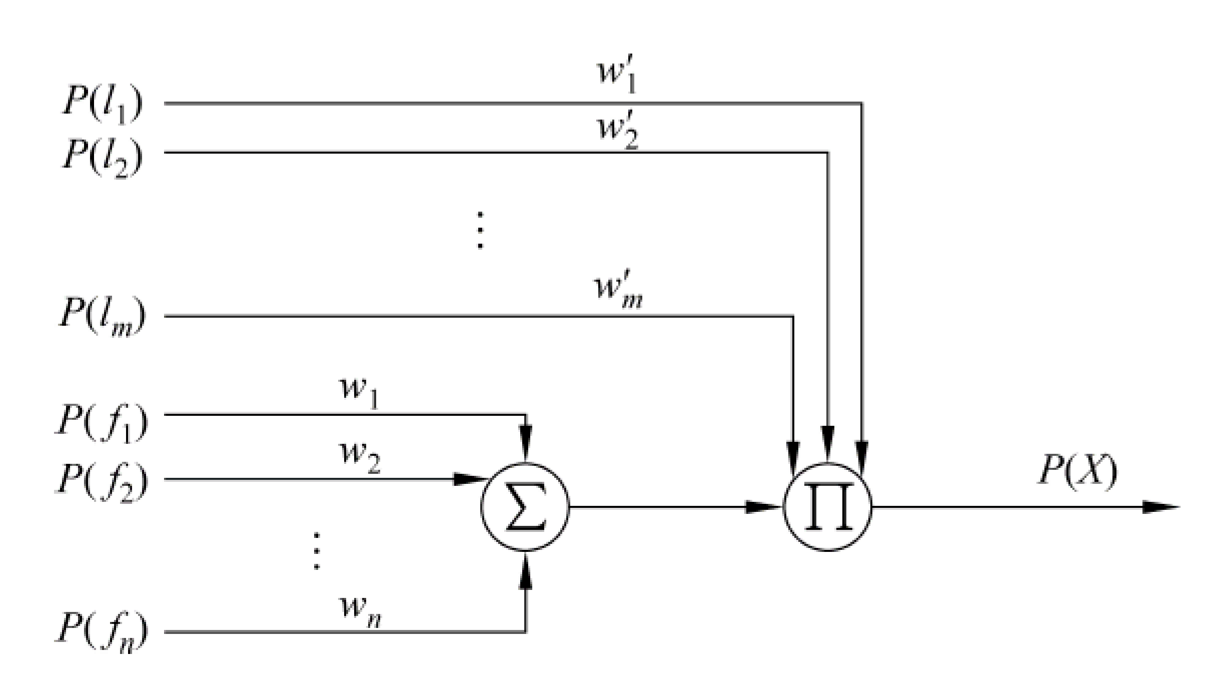



图 3.17　BLFN 模型中神经元输入耦合示意

BLFN 模型是一个由神经元构成的网络，它具有如下特点。

(1) 它采用噪声神经元模型，即每个神经元的输入和输出都是发放概率，而不是脉冲值。

(2) 每个神经元可以包含两部分输入：分别是 Feeding 输入和 Linking 输入。

(3) 神经元之间的连接权反映了它们之间的统计相关性，是通过学习得到的。

(4) 神经元的输出除了受输入影响，还受到竞争的制约。

# 3.8　神经场模型

人工神经网络被认为是具有广泛非线性关系逼近机理的学习机和非线性数据关系分类器，极其重要的是它提供了一种对于非线性函数关系的逼近理论。1985 年甘利俊一提出了信息几何并用于神经计算理论的研究[14]，其基本的思想是把所有神经网络变换所组成的空间看作为流形空间，流形空间是一般欧氏空间和 Hilbert 空间的推广，它使我们可以在一个更一般化的非线性空间和非欧氏空间上建立逼近理论和拓扑校正理论，通过几何结构和拓扑结构分析来更理解神经网络模型的变换机理和学习问题。史忠植和张建从微分流形和拓扑变换的角度来分析神经网络结构的表示、系统的识别、变换机理，以此来理解神经网络在更一般信息处理系统空间中的组织结构、定位机理。我们考虑神经计算的整体结构，提出了场组织变换与场反应变换的整体结构耦合，作为人工神经网络模型信息处理、变换机理新的表示方法；引入了模块化、层次化，还有模型的可扩展性，利用流形上的对偶几何结构对它作解释，提出相应的学习算法[974]。

## 3.8.1　神经场表示

神经网络包括输入单元、隐单元和输出单元。由于有隐单元存在，隐单元被看作为不可观测量，在信息处理中是需要被估计的值。我们把神经网络的信息处理理解为非完整数据下的信息处理。我们对神经网络的变换机理给出一种新的形式化描述。神经网络变换机理包括两个过程：神经场组织过程，通过隐单元实现从输入模式的表示；神经场反应过程，实

现从表示到输出的变换(如图 3.18 所示)。这里,$X$ 是输入空间,$X\times Z$ 是表示空间,或称作为神经场表示空间,$Y$ 是输出空间。

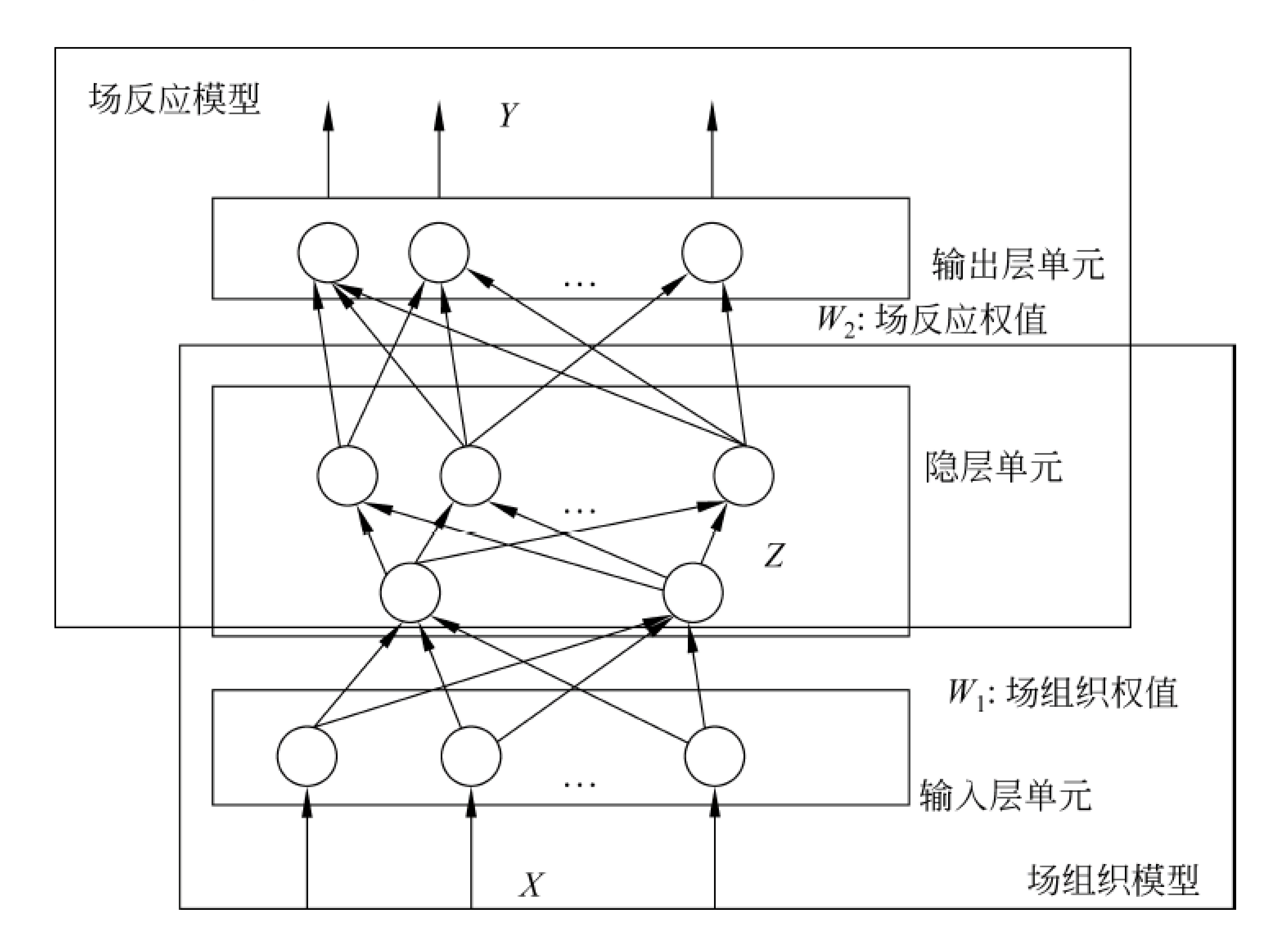



图 3.18 神经场模型

场组织变换模型表示为一个变换 $T: X\rightarrow X\times Z$,其中,$Z$ 是不可观测量空间,或是辅助信息量空间,场组织模型实现输入模式的特征表示,它可能由算法或是映射而实现;它的目的是抽取输入模式的特征结构,使期望的输出模式更容易;在神经网络中,由场组织模型所包含的连接权值,称为场组织权值或认知权值。在场组织模型权值为 $W_1$ 的情形下,场组织模型表示为

$$q_z(x)=z(x;W_1) \tag{3.61}$$

场反应模型表示一个变换 $g: X\times Z\rightarrow Y$,表示从表示空间到输出空间的变换,即通过对输入模式特征的估计后,由特征模式表示下的输出。由场反应模型所包含的神经网络连接权值,我们称之为场反应权值或产生权值。在场反应权值为 $W_2$ 的情形下,场反应模型表示为

$$p(y)=\sum_z p(z)p(y\mid z;W_2) \tag{3.62}$$

完整的人工神经网络系统形式化为含隐变量的概率分布 $P(y,z(x,W_1)|x;\theta(W_2))$,其中,$W_1$,$W_2$ 分别表示神经网络中的场组织权值和场反应权值,$\theta$ 是用来识别统计模型的参数。下面,用统计流形的理论来研究人工神经网络模型 $P(y,z(x,W_1)|x;\theta(W_2))$的结构,作为信息处理的假设,我们把人工神经网络系统嵌入到一个相关的统计流形结构中,称所有概率分布 $P(y,z|x;\theta)$组成的流形 $S$ 为人工神经网络模型 $P(y,z(x,W_1)|x;\theta(W_2))$ 的相关信息系统

$$S=P(y,z\mid x;\theta) \tag{3.63}$$

一般假设 $S$ 是一个可编码成指数簇结构的平坦流形结构,而神经网络可编码成 $S$ 的子流形结构:

$$\begin{aligned}&P(y,z(W_1)\mid x;\theta(W_2))\\&=\exp\{\Sigma r_i(y,z(W_1))\theta_i(W_2)+k(r)-\psi(\theta)\}\end{aligned} \tag{3.64}$$

$S$ 中的概率分布可由 $e$-坐标 $\theta$ 和 $m$-坐标 $\eta$ 进行识别，即

$$S = \{\theta \mid P(r(y,z) \mid x;\theta)\} = \{\eta \mid P(r(y,z) \mid x;\eta)\} \tag{3.65}$$

在对偶流形中，神经网络的场组织模型结构表示为

$$R = \{\eta(W_1) \mid P(r(y,z(x,W_1)) \mid x;\eta(W_1))\} \tag{3.66}$$

表示在隐变量已知情况下，即是完整数据下，模型的估计。

神经网络的反应模型表示为

$$M = \{\theta(W_2) \mid P(r(y,z(x,W_1)) \mid x;\theta(W_2))\} \tag{3.67}$$

在对偶流形中，神经网络的场组织模型和场反应模型变换分别表示为

$$\text{场组织变换：} T\text{：} x \rightarrow \eta(y,z(x,W_1)) \tag{3.68}$$

$$\text{场反应变换：} g\text{：} \eta(y,z(x,W_1)) \rightarrow \theta(W_2) \tag{3.69}$$

复杂结构神经网络模型可以处理简单输入/输出关系的逼近问题，但是由于复杂模型可变量（自由度）的增加，所得出的结果不一定会反映问题本身的结构，复杂模型处理简单问题泛化能力差。因此根据 occarm razor 原理，在有冗余参数下的模型结构中，应该寻找最简模式。在神经流形结构中，数据子流形 $D$ 和产生模型子流形 $M$ 的拓扑结构由单复形进行编码，可以通过单复形的边缘链分解，实现模型的分解和组织。通过边缘链分解的分维组织有非常明确的几何直观解释，对于冗余参数结构的模型，通过边缘链分解（边缘投影），消去与问题无关的冗余参数，边缘投影算子可以通过同伦变换而实现。

有了神经网络模型整体结构的流形和拓扑结构表示，神经网络对于复杂系统的信息处理和模型化可以描述为在非完整性数据估计与冗余参数结构下参数选择耦合的过程，对复杂系统模型化神经网络的组织过程描述为如图 3.19 所示的链结构。

$$\begin{array}{ccccccc}
(X^{(r)},Y) & \xrightarrow{\partial_r} & (X^{(r-1)},Y) & \xrightarrow{\partial_{r-1}}\cdots\xrightarrow{\partial_{k+1}} & (X^{(k)},Y) & \xrightarrow{\partial_k}\cdots\xrightarrow{\partial_1} & (X^{(0)},Y) \\
\downarrow T_r & & \downarrow T_{(r-1)} & & \downarrow T_k & & \downarrow T_0 \\
T_r(D^{(r)}) & \xrightarrow{\partial'_r} & T_{(r-1)}(D^{(r-1)}) & \xrightarrow{\partial'_{r-1}}\cdots\xrightarrow{\partial'_{k+1}} & T_{(k)}(D^{(k)}) & \xrightarrow{\partial_k}\cdots\xrightarrow{\partial_1} & T_0(D^0) \\
\downarrow g'_r & & \downarrow g'_{r-1} & & \downarrow g'_k & & \downarrow g'_0 \\
N^{(r)} & \xrightarrow{\partial''_r} & N^{(r-1)} & \xrightarrow{\partial''_{r-1}}\cdots\xrightarrow{\partial'''_{k+1}} & N^{(k)} & \xrightarrow{\partial''_k}\cdots\xrightarrow{\partial''_1} & N^{(0)}
\end{array}$$

图 3.19　神经场自适应神经网络模型结构

图 3.19 中，$T_i(r^i) = T_i(r^i(X,Z))$，$\partial$是边缘链分解投影变换，$T$ 是拓扑校正逼近变换，$g$ 是对偶投影变换。通过图 3.19，我们提供了一个基于结构分析的框架来讨论神经网络的分维组织和信息处理机理，它对复杂系统的模型化是通过多集模型的神经网络通过层次化的组织而实现，复杂系统的信息处理理解为在非完整数据估计和冗余参数结构下的逼近理论。以上的结构具有表达模型的可扩展性，信息处理的分维、简单模型嵌入复杂模型的机理。层次化表达为如下两个过程。

(1) 定性分析（拓扑校正逼近变换 $T$）：对输入模式的结构进行定性分析，把它嵌入相关的模型结构中，可以引入适当的隐变量和辅助量，对输入模式进行估计，这是基于结构逼近的基础。

(2) 定量分析（信息几何框架下的投影变换 $g$）：在拓扑变换的指导下，对输入模式的特征表示进行定量分析，选择优化的输出。

## 3.8.2 神经场学习理论

人工神经网络最重要的特征是学习能力,通过神经网络内部权植的更替,学习外部环境结构的特征和表示。神经网络的学习主要涉及神经网络拓扑结构的建立和设计,在固定拓扑结构的意义下,网络权值(参数)的优化和选择。一般地,神经网络的学习主要是优化某种行为的度量和测度,而这个度量和测度代表了学习问题的拓扑结构特性和模型特征。神经网络可以被看成一种学习机,如图3.20所示。

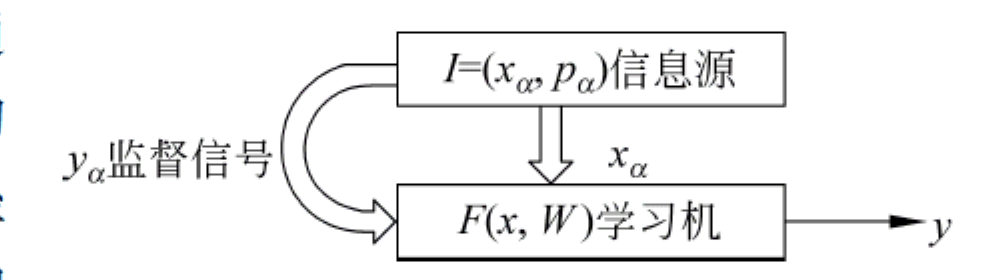



图3.20 神经网络的简单学习过程

### 1. EM学习算法

EM(Expectation-Maximization)学习算法是由登朴斯特(A P Dempster)等在1977年提出的统计模型参数估计的一种算法框架,通过交替搜索方法,使样本集的邻域函数最大化[180]。当可观测的数据被表示为 $x$ 时,我们认为它是非完整的数据,拥有一个完整的数据向量 $y$,这个数据包括损失的变量 $z$;我们识别一个概率统计模型,应该确定在完整数据下的概率分布 $P(x,z;\theta)$。完整数据的邻域函数表示为 $lc(\theta,y)=\log P(x,z;\theta)$,而非完整数据的邻域函数表示为 $l(\theta,x)=\log P(x,\theta)$。

*EM* 算法的每一次交替包括两个步骤:估计 *E* 步和最大化 *M* 步。*M* 步包括对于在一次交替过程中每一个 *E* 步所定义的邻域函数的最大化。具体 *EM* 算法的步骤分为以下的两步。

(1) *E* 步:给予可观测数据 $x$ 和由目前模型参数量 $\hat{\theta}^k$ 表示的模型,计算完整数据邻域的数学期望,得到

$$\begin{aligned} Q(\theta,\hat{\theta}^k) &= E_z[l_c(\theta \mid x)] \\ &= \int P(z \mid x,\hat{\theta}^k)\log P(x,z \mid \theta)\mathrm{d}z \end{aligned} \tag{3.70}$$

其中,$E$ 是数学期望,$\theta$ 是参数向量,是指对于非完整数据的一个估计,也就是说它是对隐变量的数学期望或完整数据的一个估计。

(2) *M* 步:确定更替值 $\hat{\theta}^{k+1}$,它计算

$$\hat{\theta}^{k+1} = \arg_\theta \mathrm{Max}(\theta,\hat{\theta}^k) \tag{3.71}$$

在一些情况下,有时很难计算 $Q(\theta,\hat{\theta}^k)$,这是由 $M$ 步所要求的过程。我们仅仅能计算 $\hat{\theta}^k$ 的一个新的值使 $Q$ 的值增加,这样的情况下,我们称之为广义的 *EM* 学习算法,参数值的变换满足

$$\hat{\theta}^{k+1} = M(\hat{\theta}^k) \tag{3.72}$$

这里

$$Q(M(\hat{\theta}^k),\hat{\theta}^k) \geqslant Q(\hat{\theta}^k,\hat{\theta}^k)$$

以上的方法,提供了一种特殊的求解问题的方法。*EM* 算法的一个交替的步骤是选择一个参数值来增加 $Q$ 的值,完整邻域的期望。

### 2. 对偶校正学习算法

在信息几何中，Amari 提出了 em 算法[15]，把神经网络模型表示成为下述的指数簇结构

$$P(y,z \mid x;\theta) = P(y,z \mid x;\theta(W_2)) \tag{3.73}$$

把这个模型嵌入到下面的指数簇结构中

$$P(y,z \mid x;\theta(W_2)) = \exp\{\Sigma r_i(y,z)\theta_1(W_2) + k(r) - \psi(\theta(W_2))\} \tag{3.74}$$

我们把神经网络的权值分离成场组织权值和场反应权值，神经网络的模型被描述为

$$P(y,z \mid x;\theta) = P(y,z(W_1) \mid x;\theta(W_2)) \tag{3.75}$$

把这个模型嵌入到下面的指数簇结构中

$$\begin{aligned} &P(y,z(W_1) \mid x;\theta(W_2)) \\ &= \exp\{\Sigma r_i(y,z(x;W_1))\theta_i(W_2) + k(r(y,z(W_1))) - \psi(\theta\ (W_2))\} \end{aligned} \tag{3.76}$$

在训练样本集 $Y=\{x^t,y^t\},t=1,2,\cdots,N\}$ 下，神经网络模型的信息几何学习理论，称为对偶校正学习算法(DCL)。神经网络包括两类连接权值，场组织模型连接权值 $W_1$；由场组织模型作用产生表示 $z(W_1)$，称为非完整数据结构的估计；同样的道理，神经网络的场反应变换连接权值 $W_2$，当表示为 $z$，由产生权值作用所产生的输出模式为 $y$，神经网络模型表示为概率分布 $P(y,z(W_1)|x;\ \theta(W_2))$。

对偶校正学习算法是基于非完整数据假设的微分流形投影理论的一种方法，其基本的结构依赖于神经网络模型的指数簇结构编码。当神经网络模型 $M$ 作为子流形嵌入到带隐变量的指数簇流形 $S$ 中，由 $\theta$ 坐标识别；训练集 $Y=\{x^t,y^t\},t=1,2,\cdots,N\}$ 被看作是非完整数据，$z$ 是不可观测量，即 $z$ 可任意取值，这样对于训练集 $Y$，形成 $S$ 中的一个子流形 $D$，称为数据子流形，由 $\eta$ 坐标识别。对于输入样本$(x_i,y_i)$，可以得到数据子流形 $D_i$

$$D_i = \{W_1 \mid P(r(y_i,z(W_1))) \mid x_i;\eta(W_1)\}\} \tag{3.77}$$

同样可以引入产生模型子流形 $M$，

$$M_i = \{W_2 \mid P(r(y_i,z) \mid x_i;\theta(W_2))\} \tag{3.78}$$

从信息几何的观点来看，神经网络的学习算法的优化目标是减小 $D_i$ 和 $M_i$ 的 Kullback-Leibler 差异度量

$$K_i(D_i,M_i) = \mathrm{Min}_{W_1\in D_i,W_2\in M_i} K_i(W_1,W_2) \tag{3.79}$$

在训练样本集$\{(x_i,y_i),i=1,2,\cdots,T\}$的情况下，学习被描述为优化下式的过程：

$$K(D,M) = \sum_{i=1}^{T} k_i(D_i,M_i) \tag{3.80}$$

$D$ 和 $M$ 的 Kullback-Leibler 熵减小要通过对偶校正学习算法实现。

对偶校正学习算法 DCL 有如下步骤。

(1) 初始化：取神经网络初值，对 $W_2$ 进行随机的初始化赋值为 $W_2^0$，从而确定 $P^0\in M$。

(2) 从第 $i$ 步起，$i=0,1,\cdots$，重复以下的过程。

① $E$ 步：沿 $e$-测地线投影 $P_i$ 到 $D$，得到投影点 $Q_i\in D$，减小 Kullback-Leibler 差异度量 $K(Q,P_i)$，$Q\in D$。

② $M$ 步：沿 $m$-测地线投影 $Q_i$ 到 $M$，得到投影点 $P_{i+1}\in M$，减小 Kullback-Leibler 差异度量 $K(Q_i,P)$，$P\in M$。

### 3. 基于拓扑结构逼近的学习理论

在神经场理论的框架下基于拓扑结构逼近的学习算法理论，基本的思想可表示为人的感知系统模型，用隐变量来估计和发现输入模式的主要结构，形成输入模式相关的问题结构表示，通过这个结构表示来指出相关的结构估计，表示空间的结构是一个中间结构，它有助于得到期望的结果。这一部分，我们发展了基于代数拓扑概念的结构表示，用拓扑结构表示作为人工神经网络结构抽取表示的中间过程，对于未知系统的逼近密切关系着表示空间中的中间过程。这个表示空间与 PCA、LVQ SOM 和 MDL 的表示机理不同，也不同于 Helmhotz 机的描述，它突出由代数拓扑中单、复形结构的组织空间作为表示空间，对于神经网络数值结构的表示和编码机理。神经网络变换机理包括两个变过程，一是抽取输入的拓扑表示，形成神经场表示；另一个是在表示下期望值的输出。与之相适应的学习过程也包括两个方面，抽取输入模式特征和最佳结构的估计以满足期望的输出。

我们把人的感知系统看作为统计学习机，用流形上的拓扑概念来发展学习理论，一个输入为 $x$ 输出为 $y$ 的未知目标系统，我们用 $q(y|x)$ 来表示，一个基本的问题就是怎样来描述目标系统的结构问题，把它嵌入到一个什么样的信息处理结构中，与问题相关的系统的结构我们用统计流形学习机

$$S = \{P(y,z \mid x,\Theta)\} \tag{3.81}$$

来进行表示，这里 $x,y$ 是可观测数据，$z$ 是辅助数据量或是不可观测数据，也就是在非完整信息处理的框架下，对偶统计流形作为系统的结构描述方法。作为对偶流形，可以在 $S$ 上引入对偶坐标 $\theta$ 和 $\eta$，即用两种坐标来编码系统，来对系统进行识别；$S$ 可表示为

$$S = \{\theta \mid P(y,z \mid x;\theta)\} = \{\eta \mid P(y,z \mid x;\eta)\} \tag{3.82}$$

$P(y,z|x,\Theta)$ 可由对偶坐标 $\theta$ 和 $\eta$ 来进行表示和识别。

在神经场理论中，采用代数拓扑中的单复形结构来表示空间。单复形结构是模型结构嵌入最直观的拓扑概念，为研究简单模型向复杂模型的嵌入给出了非常直观的形象。在非完整数据信息处理的框架下，假设训练样本集 $(x_i,y_i)$，$i=\{1,2,\cdots,T\}$ 所载信息量表示为 $t(x,y,z)$，$t$ 是 $(x,y,z)$ 的函数，由于 $z$ 为辅助量或不可观测量，当一个辅助量 $z$ 确定，就可看作为一个完整信息，从而决定一个系统，用 $\eta$ 坐标来识别系统，当 $z$ 变化时，形成一个数据子流形，为

$$D = \{\eta \mid P(t(x,y,z) \mid x;\eta)\} \tag{3.83}$$

$D$ 是 $S$ 的一个子流形，下面为了研究对于输入信息量结构的描述，我们引进下述的 $D$ 的拓扑结构编码表示，即 $D$ 的表示空间为 $V$。

$$V = \left\{\eta \in S \mid \eta = \sum_{i=0}^{r} c_i \eta_{qi},\ \sum_{i=0}^{r} c_i = 1, c_i \geqslant 0\right\} \tag{3.84}$$

这里，$q_i$ 是 $S$ 上 $r$ 个点。实际上，$V$ 是由流形 $S$ 上 $r$ 个数据点张成的平坦流形，$D$ 是 $V$ 的弯曲子流形，$V$ 是 $\eta$ 平坦子流形结构；从拓扑的结构看，$D$ 和 $V$ 有相同的拓扑结构。引入了表示空间 $V$，则数据子流形的拓扑结构可由 $V$ 的结构进行表示。数据子流形 $D$ 作为 $V$ 的弯曲子流嵌入 $V$ 中。同样也可以进一步地用复形结构表示输入模式的结构。表示空间 $V$ 称为复形结构编码的表示空间。

可分子流形结构编码：假设在 $\eta$ 坐标编码中，可观测变量与隐变量是可分离的，则数据

子流形 $D$ 的拓扑结构被编码成为 $V$

$$V = \{\eta = (\eta_v, \eta_h) \mid P(t(x,y,z) \mid x;\eta)\} \tag{3.85}$$

这是最为简单的结构编码。

和数据子流形的结构编码一样，神经网络模型产生结构可以由 $\theta$ 坐标进行表示，形成神经网络子流形 $M$，$M$ 的拓扑结构也可以编码成 $S$ 上的 $\theta$-复形结构，$M$ 的拓扑结构编码为

$$N = \left\{\theta \in S \mid \theta = \sum_{i=0}^{r} c_i \theta_{pi}, \sum_{i=0}^{r} c_i = 1, c_i \geqslant 0\right\} \tag{3.86}$$

$N$ 是 $S$ 上的 $\theta$-平坦子流形结构，而 $M$ 是 $S$ 的弯曲子流形。$N$ 可以看作为 $M$ 的主要结构，$N$ 和 $M$ 的相同的拓扑不变结构。从拓扑不变量的角度看，$M$ 和 $N$ 是同伦不变的，即 $M$ 可以通过同伦变换收缩到 $N$。$N$ 实际上也是对于 $M$ 结构定性描述。

数据子流形是对于输入数据模型结构的描述，是对认知模型结构的认识，由神经网络的认知权值结构作用而产生；而神经网络模型是对产生式结构的描述，由神经网络产生模型而形成，由产生权值作用而产生。以上两个方面可以看作为同一个过程认识的多集模式表示，即一个问题由两种方式表示，从形式上说，认知模型表示为相关系统 $S$ 子流形 $D$ 和对应的拓扑结构编码 $V$：

$$D = \{W_1 \mid P(y, z(W_1) \mid x; \eta(W_1))\} \tag{3.87}$$

$$V = \left\{\eta \in S \mid \eta = \sum_{i=0}^{r} c_i \eta_{qi}, \quad \sum_{i=0}^{r} c_i = 1, c_i \geqslant 0\right\} \tag{3.88}$$

产生模型表示为 $S$ 的子流形 $M$ 和对应的拓扑结构编码 $N$。

$$D = \{W_2 \mid P(y, z \mid x; \theta(W_2))\} \tag{3.89}$$

$$V = \left\{\eta \in S \mid \eta = \sum_{i=0}^{S} c_i \eta_{pi}, \quad \sum_{i=0}^{S} c_i = 1, c_i \geqslant 0\right\} \tag{3.90}$$

作为同一个问题的不同表示方法，学习的问题就是两种表示的耦合问题，两种表示拓扑结构的定性逼近和在此基础指导下的定量分析逼近。

我们看到对于数据子流形 $D$，可以形成 $D$ 的拓扑结构表示 $V$，$V$ 是 $D$ 的拓扑结构编码；同样对于产生式神经网络子流形 $M$，也有一个单复形结构拓扑编码 $N$ 作为其结构的描述。学习的策略包括两个方面，一个是使数据子流形的拓扑结构 $V$，与 $N$ 的结构尽可能一致，即通过拓扑特征映射 $T$，减小 $V$ 与 $N$ 间的差异

$$T: V \to N$$

这个拓扑变换，将实现拓扑特征抽取，把神经场的表示变换到神经网络产生式模型的决结构中，实现结构的逼近；另一个过程是对偶校正学习算法，在拓扑校正后，通过两个表示间交替地减小 Kullback-Leiber 差异度量，即通过对偶学习算法，减小 $T(D)$ 与 $M$ 间的 Kullback-Leibler 差异度量。更具体地说，基于拓扑结构逼近的学习算法 TAC 包括如下两个过程。

(1) TAC 拓扑变换：在前面看到，拓扑特征变换可由变换算子实现或由算法实现。在结构编码的情况下，我们用变换算子来实现拓扑变换。假设 $V$ 和 $N$ 的顶点满足关系 $K(q_1, p_1) \leqslant K(q_2, p_2) \leqslant \cdots \leqslant K(q_r, p_r)$ 且 $r \leqslant s$，则

$$T: V \to N, \quad \eta = \sum_{i=0}^{r} c_i \eta_{qi} \to \sum_{i=0}^{r} c_i \theta_{pi} \tag{3.91}$$

这里，$K$ 是 Kullback-Leibler 差异度量

$$E_{\text{topo}}(V,N)=\sum_{i=0}^{r}K(q_i,p_i) \tag{3.92}$$

称为拓扑信息熵,经过 $T$ 变换我们看到拓扑信息熵为0。

(2) 对偶校正学习算法(DCL):经过拓扑校正映射 $T$ 的作用后,形成了 $D$ 与 $M$ 间拓扑结构的校正,即 $T(D)$ 与 $M$ 间有相同的拓扑结构;下一步,即是在相同的拓扑指导下,运用信息几何的学习理论(对偶校正学习算法或 em 学习算法)交替减小 $T(D)$ 与 $M$ 间的 Kullback-Leibler 差异度量,学习的优化目标为

$$K(W_1,W_2)=\text{Min}K(T(D),M)=\text{Min}_{Q\in T(D(W_1)),P\in M(W_2))}K(Q,P) \tag{3.93}$$

上述过程可用信息几何的 em 学习算法实现,描述为

① $e$-投影步:固定 $W_2$,得到 $P$,$e$-投影 $P$ 到 $T(D)$,从而得到 $T(Q(W_2)\in T(D)$,找出 $W_1$,减小 $K(W_1,W_2)$。

② $m$-投影步:固定 $W_1$,得到 $Q$,$m$-投影 $T(Q)$ 到 $M$,从而得到 $P\in M(W_2)$,找出 $W_2$ 减小 $K(W_1,W_2)$。

## 3.9 超限学习机

单隐层前馈神经网络(Single-hidden Layer Feedforward Neural Network,SLFN)之所以能够在很多领域得到广泛应用,是因为它有很多优点:①具有很强的学习能力,能够逼近复杂非线性函数;②能够解决传统参数方法无法解决的问题。但另一方面缺乏快速学习方法,也使其很多时候无法满足实际需要。产生这种情况的主要原因是:①传统的误差反向传播方法主要基于梯度下降的思想,需要多次迭代;②网络的所有参数都需要在训练过程中迭代确定。

黄广斌等研究了有限集情况下 SLFN 的学习能力,只和隐层节点的数目有关,而和输入层的权值无关。在此基础上,黄广斌提出了超限学习机(Extreme Learning Machine,ELM)[335],设置合适的隐层节点数,为输入权和隐层偏差进行随机赋值,然后输出层权值通过最小二乘法得到。整个过程一次完成,无须迭代,与 BP 相比速度显著提高。超限学习机的结构如图3.21所示。

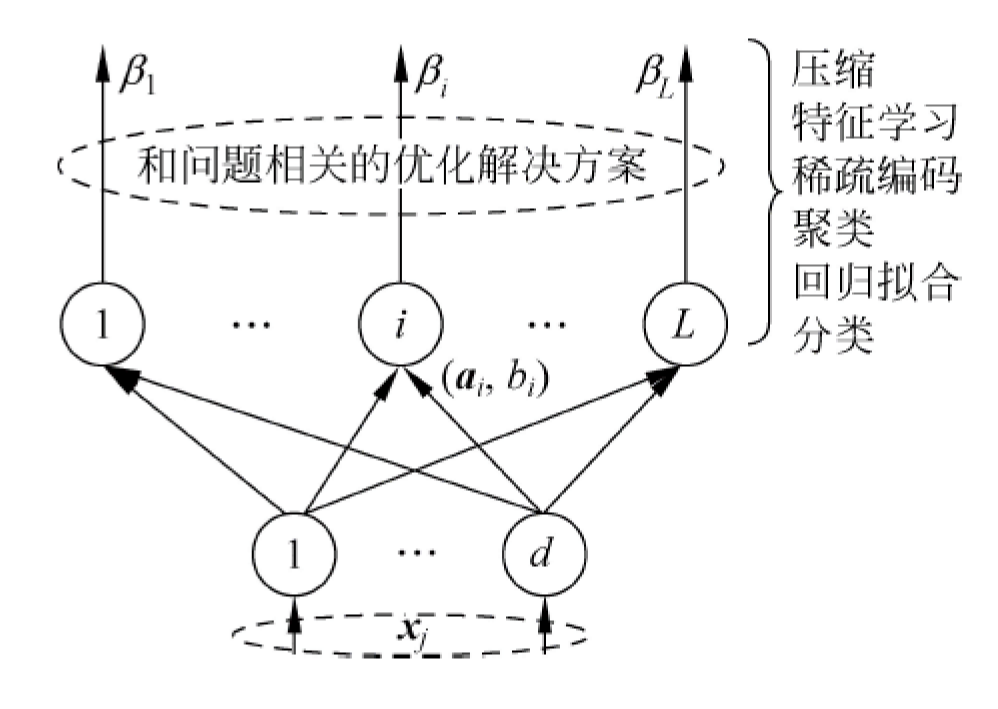



图3.21 超限学习机的结构

图3.21中,超限学习机的网路输出:

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, \boldsymbol{X}) \tag{3.94}$$

超限学习机的特征映射：

$$\boldsymbol{h}(x) = [G(\boldsymbol{a}_1, b_1, \boldsymbol{x}), \cdots, G(\boldsymbol{a}_L, b_L, \boldsymbol{x})] \tag{3.95}$$

隐节点的输出函数

$$\begin{aligned} &\text{Sigmoid:} && G(\boldsymbol{a}_i, b_i, \boldsymbol{x}) = g(\boldsymbol{a}_i \cdot \boldsymbol{x} + b_i) \\ &\text{RBF:} && G(\boldsymbol{a}_i, b_i, \boldsymbol{x}) = g(b_i \| \boldsymbol{x} - \boldsymbol{a}_i \|) \\ &\text{Fourier Series:} && G(\boldsymbol{a}_i, b_i, \boldsymbol{x}) = \cos(\boldsymbol{a}_i \cdot \boldsymbol{x} + b_i) \end{aligned} \tag{3.96}$$

## 3.10　功能柱神经网络模型

自 1957 年莫特卡斯勒(V B Mountcastle)发现功能柱结构以来，已有许多研究结果表明，在不同物种(鼠、猫、兔、猴和人等)的视皮层、听皮层、体感皮层、运动皮层，以及其他联合皮层当中都存在功能柱结构。这些结果表明，功能柱是皮层中一个普遍的结构，是结构和生理上的基本单元，这些柱的活动构成了整个大脑皮层活动的基础[517]。

为了深刻地理解功能柱的生物学意义和在信息加工中所起的作用，研究者开展了许多数学建模研究。模型研究中最常见的是采用 Wilson-Cowan 方程来描述功能柱，例如：舒斯特(H G Shuster)等人模拟视皮层中发现的同步振荡现象[664]；詹森(B H Jansen)等人提出了耦合功能柱模型产生了类 EEG 波形和诱发电位[342]；富凯(T Fukai)设计了功能柱式的网络模型来模拟视觉图样的获取等[250]。另有一些功能柱模型是描述功能柱振荡活动的相位模型。只有少数模型是基于单神经元的，如：弗朗森(E Fransén)等人把传统网络中的单细胞代换成多细胞构成的功能柱，构建了一个吸引子网络，来模拟工作记忆；汉塞勒(D Hansel)等人根据视皮层朝向柱的结构构建了一个超柱模型，研究其中的同步性和混沌特性，并对朝向选择性的功能柱机理做出解释。

李速等采用模型神经元作为基本单元，按皮层功能结构组织起来的功能柱模型，探索这些发放模式与外界输入和网络结构的关系，研究多个功能柱联结起来组成的网络模型，在活动模式上有什么新的特点[879]。

### 3.10.1　模型与方法

#### 1. 神经元模型

希望从发放的单神经元出发来构建功能柱模型，因此选用了 Rose-Hindmarsh 方程来描述单神经元：

$$\begin{cases} \dot{x} = y + ax^3 - bx^2 - z + I_{\text{syn}} + I_{\text{stim}} \\ \dot{y} = c - dx^2 - y \\ \dot{z} = r[s(x - x_0) - z] \end{cases} \tag{3.97}$$

其中，$x$ 代表膜电位，$y$ 表示快速回复电流，$z$ 描述慢变化的调整电流，$I_{\text{syn}}$ 表示突触电流，$I_{\text{stim}}$ 表示外界输入，$a,b,c,d,r,s,x_0$ 均为常数，在这里取值为 $a=1,b=3,c=1,d=5,s=2$，

$x_0=-1.6$。在 Rose-Hindmarsh 模型中的时间尺度为 5 单位相当于 1ms。

根据生理试验的结果,皮层功能柱中的神经元按生理特性分主要有两类:规则发放型细胞(Regular-Spiking,RS)和快速发放型细胞(Fast-Spiking,FS)。RS 细胞是兴奋型的,在形态上均为锥体细胞,它的特性是明显而快速的发放频率适应性,对于持续电流刺激下随着时间延长发放频率快速降低。FS 细胞是抑制型的,在形态上通常为非锥体细胞,它的生理特性是对持续电流输入的频率适应性较低。我们采用不同的参数 $r$ 来表现这两种细胞的特征:$r_{RS}=0.015$,$r_{FS}=0.001$。在图 3.22(a)当中画出了两种神经元的发放图样和频率-时间曲线。

## 2. 突触模型

模型采用基于电流的突触模型,在突触前细胞的每个动作电位都将触发突触后细胞的 $I_{syn}$ 输入。突触电流 $I_{syn}$ 表示为:

$$I_{syn}=g_{syn}V_{syn}(e^{-t/\tau_1}-e^{-t/\tau_2}) \tag{3.98}$$

其中 $g_{syn}$ 为膜电导,$\tau_1$ 和 $\tau_2$ 是时间常数,$V_{syn}$ 表示突触后电位。用 $V_{syn}$ 来调节突触耦合的强度。RS 细胞之间兴奋型连接的 $V_{syn}$ 用 $V_{RR}$ 表示,同理,从 RS 细胞投射到 FS 细胞、从 FS 细胞投射到 RS 细胞的 $V_{syn}$ 分别表示为 $V_{RF}$ 和 $V_{FR}$。参数设置为 $g_{RR}=4$,$\tau_{1(RR)}=3$,$\tau_{2(RR)}=2$,$g_{RF}=8$,$\tau_{1(RF)}=1$,$\tau_{2(RF)}=0.7$,$g_{FR}=4$,$\tau_{1(FR)}=3$,$\tau_{2(FR)}=2$。$V_{FR}$ 始终设为 $-1$。$V_{RR}$ 和 $V_{RF}$ 在模拟过程中从 0.1 到 1 之间变化。几种突触中的电流和突触后电位变化见图 3.22(b)。

## 3. 网络结构

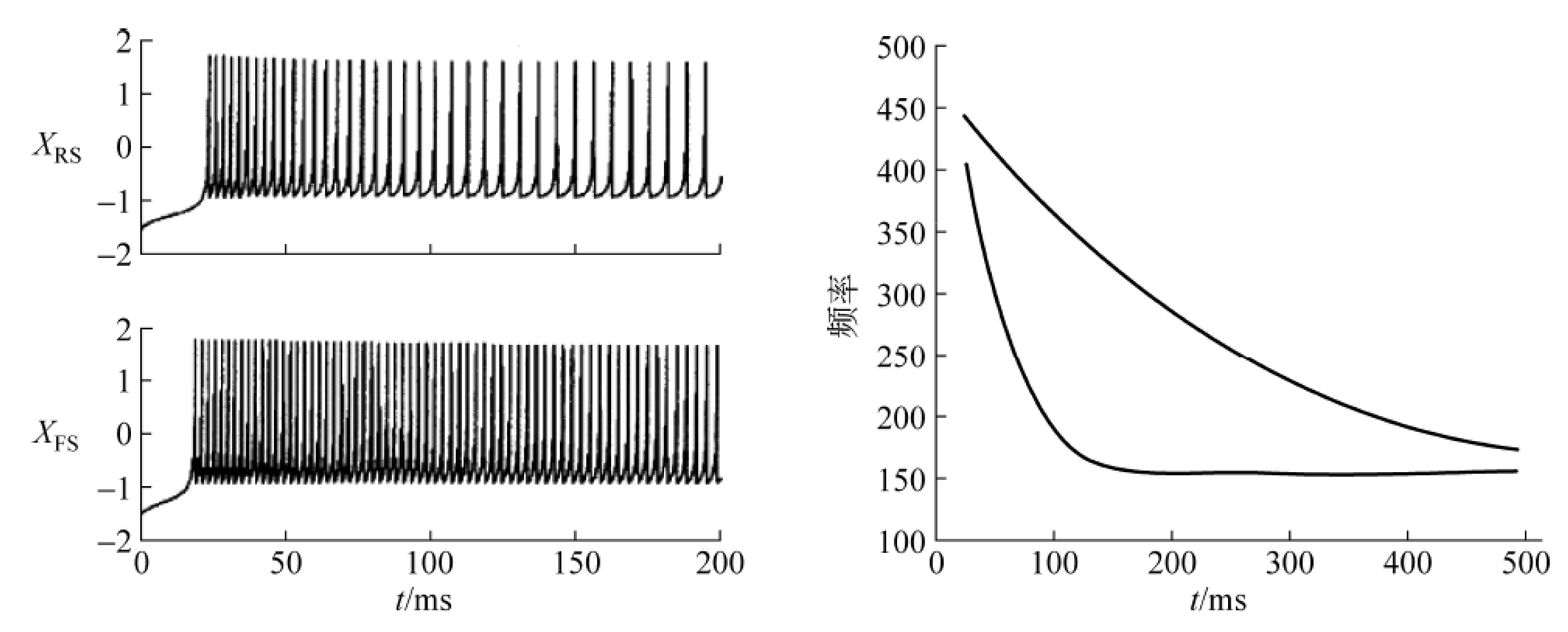
网络结构不同于以往模型的是,功能柱由发放的单神经元构成,再以功能柱为基本结构模块,构成更大型的网络。作为模块的功能柱,其内部结构是相对固定的,功能柱内部神经元连接也较为丰富,而功能柱与功能柱之间的连接则稀疏得多。

对功能柱中的神经元数量和类型做了简化,按生理资料,兴奋型神经元约占 80%,抑制型神经元约占 20%。我们的模型中,一个功能柱由 15 个神经元组成,12 个 RS 神经元,3 个 FS 神经元。

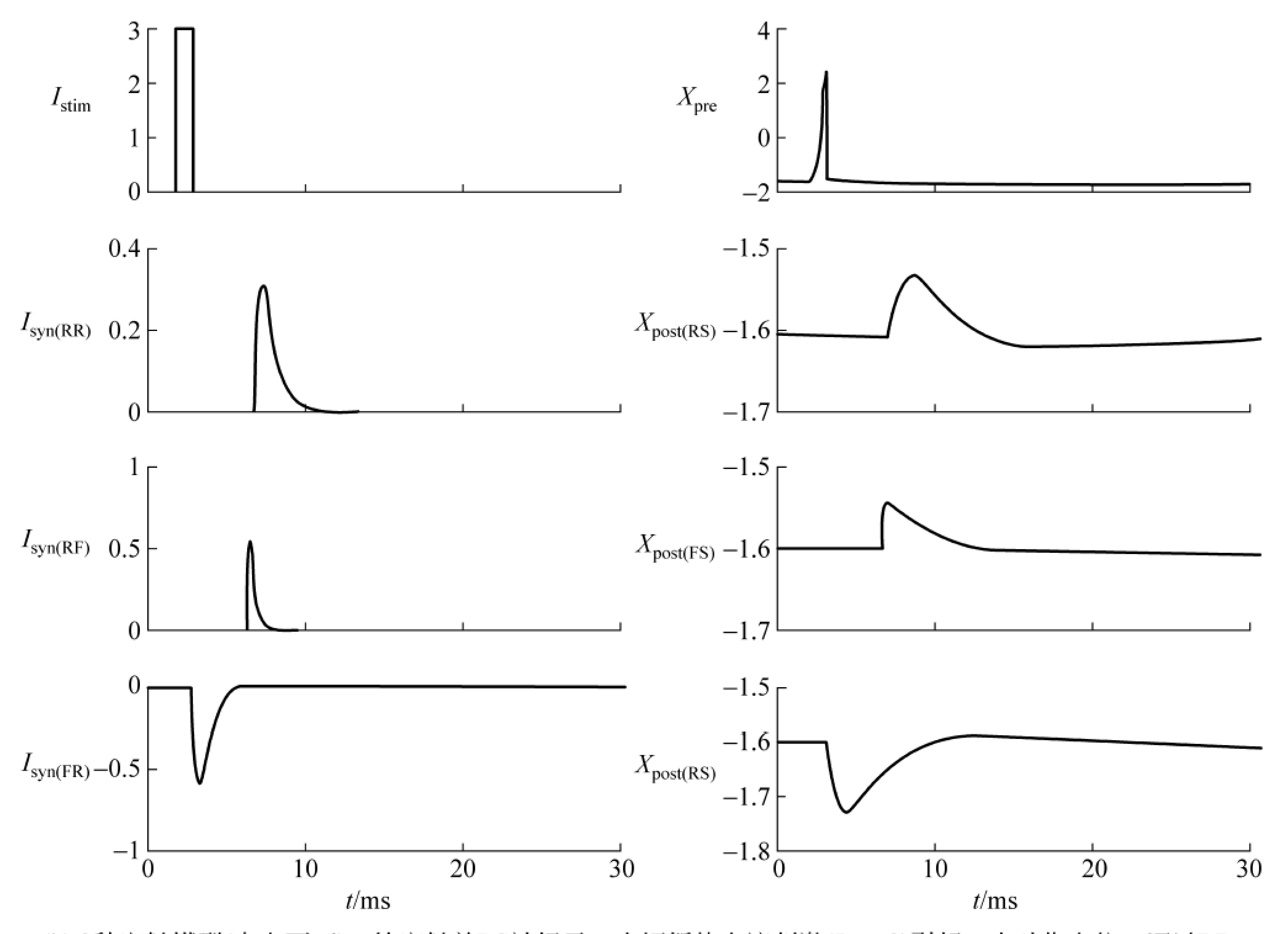
在功能柱中,锥体细胞的轴突上发出丰富的回返侧枝,投射到同一功能柱内其他锥体细胞,形成兴奋型回路。这种连接把一个功能柱内的锥体细胞耦合成为一个相互兴奋的系统。在本文的模型中,每个 RS 细胞随机连接到其他 6 个 RS 细胞。这种 RS 到 RS 的连接(以下称为 RR)的延时是 1.2 单位时间,标准差为 2.5%。

锥体细胞的轴突侧枝也终止于抑制性中间神经元,抑制性中间神经元则在锥体细胞上形成抑制性突触,这样便形成了功能柱内的抑制性反馈回路。在我们的模型中,每个 FS 神经元接受从 5 个 RS 神经元来的突触,也就是说平均每个 RS 细胞投射到 1.25 个 FS 细胞。对每个 RS 细胞而言,它发出的纤维终止于 RS 细胞和 FS 细胞的比率符合解剖学统计结果。RF 突触连接的延迟是 0.8±2.5%时间单位。

在形态学观察中,锥体细胞具有较长的轴突,而 FS 细胞往往轴突很短。我们假定抑制性的 FS 神经元只能作功能柱内的局域投射,而只有 RS 神经元能够将轴突分支投射到其他功能柱内。功能柱之间相互连接。我们假定功能柱内部的连接较强,功能柱间则连接稀疏




(a) 两种神经元模型：锥体细胞(规则发放型, RS)神经元和抑制型神经元(快速发放型, FS)的不同发放模式
左图：两种神经元在持续电流刺激下的发放模式，输入电流$I_{stim}$均开始于0时刻，大小为0.23
右图：两种神经元发放的即时频率变化曲线




(b) 3种突触模型(自上而下)：给突触前RS神经元一个短暂的电流刺激($I_{stim}$=3)引起一个动作电位，通过RR连接、RF连接中的突触电流($I_{syn}$)，在突触后的一个RS细胞和FS细胞中分别引起了膜电位($x$)变化。最后一图则是表示FR型突触电流和突触后RS细胞的膜电位变化(突触前FS神经元未画出)

图 3.22　神经元和突触模型

而且耦合度弱。在目前的网络模型中，我们只设定了一种功能柱之间的连接，即从一个功能柱内的 RS 神经元连接到目标柱内的 RS 神经元，形成功能柱之间的兴奋性连接(以下称为 iRR，符号前标注 $i$ 表示功能柱之间，下同)。这种连接的数量是，2 个突触前细胞投射到 6 个突触后细胞。可以看出，相对于柱内连接，柱间的连接是非常稀疏的。连接参数为 $g_{iRR}=4$，$\tau_{1(iRR)}=3$，$\tau_{2(iRR)}=2$，延迟是 2±2.5%时间单位。

### 4. 网络的输入

网络的输入是随机的脉冲输入，模拟功能柱接受的来自丘脑和皮层其他区域来的脉冲刺激。外界输入只刺激 12 个 RS 细胞，每个细胞的脉冲序列是随机的，并在各细胞间相互独立。每个脉冲都在受刺激的细胞中触发 $I_{stim}$ 电流输入，其方程如下：

$$I_{stim} = g_{stim}V_{stim}(e^{-t/\tau_1} - e^{-t/\tau_2}) \tag{3.99}$$

其中，$g_{stim}=4$，$V_{stim}=0.08$，$\tau_{1(stim)}=3$，$\tau_{2(stim)}=2$。可以看出外界刺激的幅度比功能柱内部的突触耦合强度要小得多。

外界输入的强度定义为每个 RS 细胞在每个时间间隔中收到一个脉冲刺激的概率 $P_{stim}$。

**注意**：时间间隔和 $P_{stim}$ 都很小，可以认为一个细胞在一个时间间隔内至多只能接受一个脉冲刺激。

这里采用了两种刺激方式，一种是恒定刺激，其中 $P_{stim}$ 保持不变，另一种是周期性刺激，$P_{stim}$ 定义为时间的函数：

$$P_{stim}(t) = E_{stim} + A_{stim} * \sin(\omega t \cdot 2\pi/T) \tag{3.100}$$

$E_{stim}$ 和 $A_{stim}$ 分别是 $P_{stim}$ 的均值和振幅。$T$ 等于 5000(在模型中等于 1s)，因此 $\omega$ 就是脉冲概率振荡的频率。

### 5. 度量指标和计算方法

功能柱的局部场电位 $U$ 定义为 12 个 RS 细胞膜电位的平均值。功能柱输出的功率普通过场电位自相关函数的傅里叶变换得到。

为衡量功能柱内神经元发放的同步程度，引入同步性 $\kappa$ 这一度量，它的计算基于功能柱神经元对的相关系数。把一段长的时间 $T$ 按分解为小片断 $\lambda$，每一小段内的膜电位加以平均，就得到了一个新的膜电位序列 $X_i(l)$，$l=1,2,3,\cdots,n(n=T/\lambda)$。计算神经元 $i$ 和 $j$ $(1\leqslant i,j\leqslant 12)$ 之间的相关系数 $\kappa_{ij}$：

$$\kappa_{ij} = \frac{\sum_l (x_i(l) - \overline{x_i})(x_j(l) - \overline{x_j})}{\sqrt{\sum_l (x_i(l) - \overline{x_i})^2 \sum_l (x_j(l) - \overline{x_j})^2}} \tag{3.101}$$

群体的相关性 $\kappa$ 定义为每两个 RS 细胞间 $\kappa_{ij}$ 的平均值。本文中时间片断 $\lambda$ 设为 1，即相当于实际的 0.2ms。

为避免网络设定和输入刺激的初始条件的影响，网络输出的前 200ms 舍弃，所有结果都从模拟输出 200ms 以后开始计算。由于网络连接和刺激有随机性，以下结果均是重复 3～5 次的平均值。数值计算的时间步长为 0.01ms。

## 3.10.2 单功能柱模型的模拟结果

我们首先对功能柱模型输入恒定刺激，研究功能柱内部的锥体细胞的发放的同步性和整个网络输出的节律现象。

图 3.23 显示了几种不同的节律输出。其中，图 3.23(a)是典型的 alpha 节律，其场电位波形有明显的纺锤形振荡，其振幅增大减小的周期约为 1.5s。场电位的能谱表明振荡频率

处于 alpha 频段（14～30Hz）。图 3.23(b)中显示的是处于 beta 频段（8～13Hz）的振荡。图 3.23(c)中的振荡频率为 5Hz 左右的方波，发放期较长而且振幅很大。图 3.23(d)显示的是 40Hz 振荡。在图 3.23(e)中，神经元进入持续发放状态，能谱也没有明显的峰，这种状态上下文中称为持续兴奋。以上各种节律均是由模型在不同参数条件下产生，下文中将详细讨论不同的条件对皮层节律和同步性的影响。

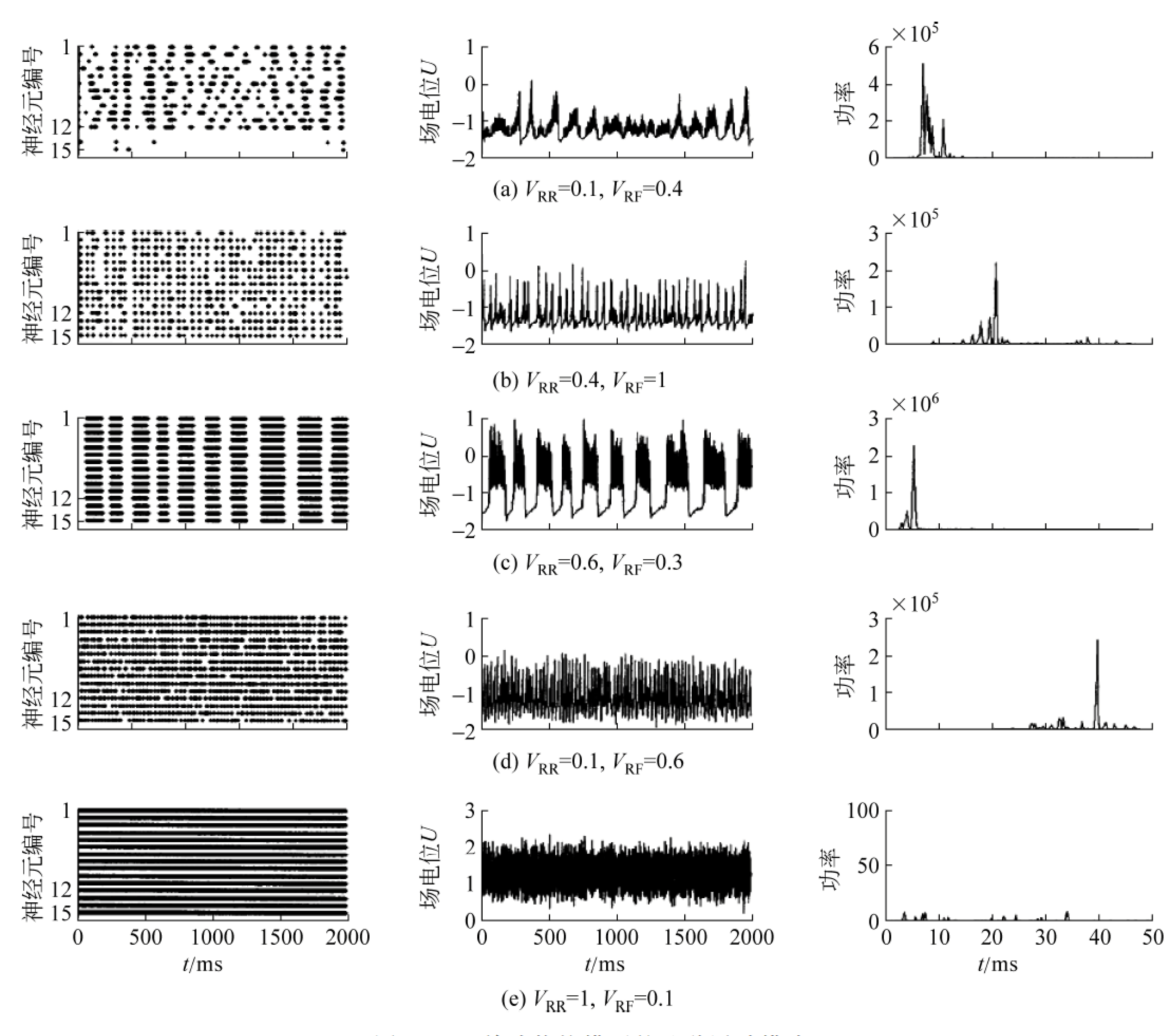



图 3.23　单功能柱模型的几种活动模式

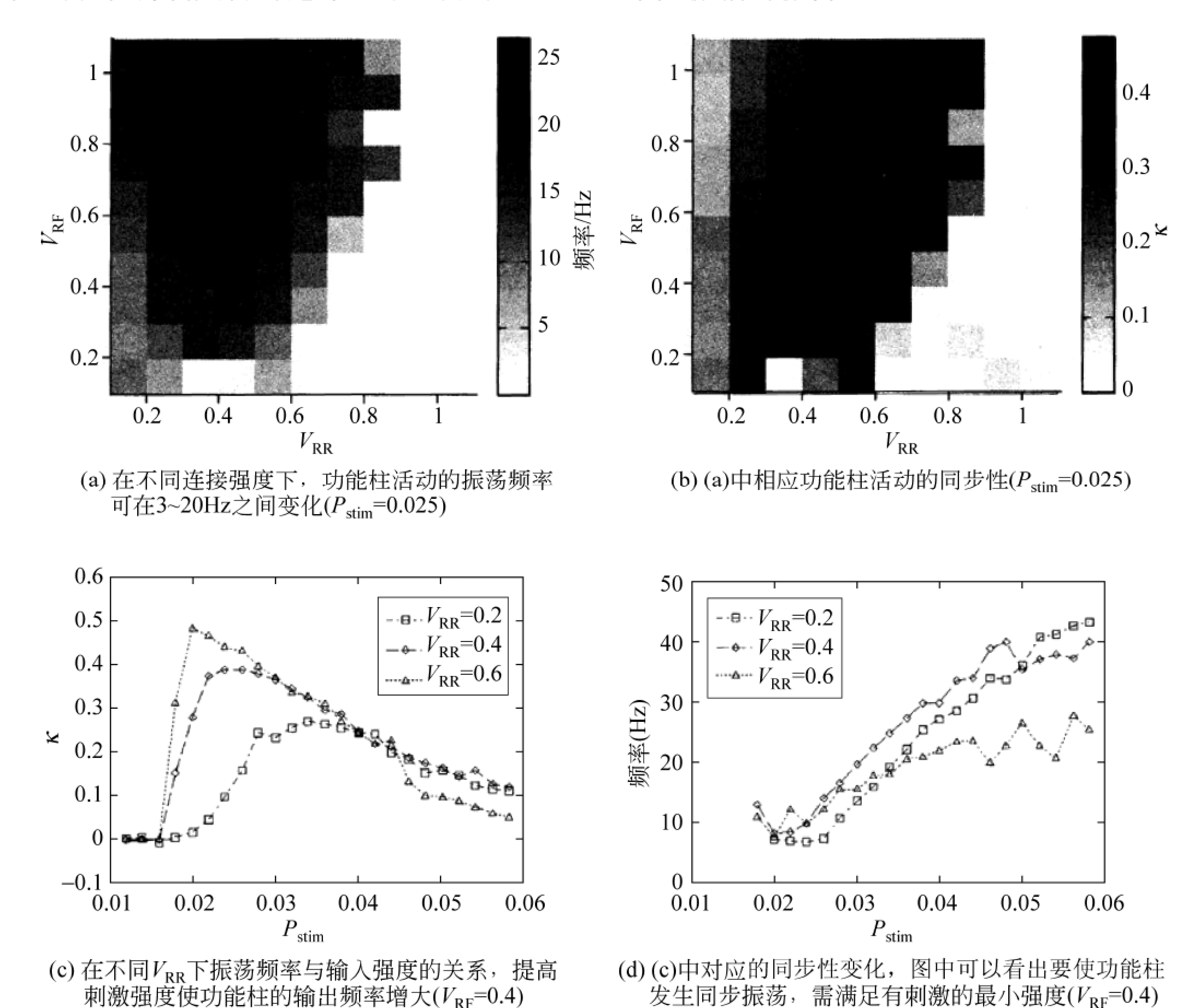
左列：神经元发放图形。纵轴 15 个神经元中，1～12 为 RS 细胞，13～15 为 FS 细胞，每个点表示一个动作电位。中列：功能柱的场电位。右列：场电位的功率谱，表示功能柱的输出频率。(a)$V_{RR}=0.1$，$V_{RF}=0.4$；(b)$V_{RR}=0.4$，$V_{RF}=1$；(c)$V_{RR}=0.6$，$V_{RF}=0.3$；(d)$V_{RR}=0.1$，$V_{RF}=0.6$；(e)$V_{RR}=1$，$V_{RF}=0.1$。除(d)图中 $P_{stim}=0.04$ 外，其余各图 $P_{stim}=0.025$。

## 1. 同步性与突触耦合强度的关系

为进一步研究同步振荡和突触耦合强度的关系，在模拟过程中改变 $V_{RR}$ 和 $V_{RF}$ 的值。$P_{stim}$ 设定为 0.025 且不变。图 3.24(a)中，网络输出的频率表示在 $V_{RR}$ 和 $V_{RF}$ 构成的二维参数空间中。随着兴奋性连接强度的增大，输出频率先增大，再减小，直到在图的右下角，即兴奋型连接强度大而抑制性连接强度小的区域，输出节律为 0，对应于图 3.23(e)中的持续兴

奋状态,此时网络过兴奋。而抑制性回路中耦合强度 $V_{RF}$ 的增强则单调地增加网络的频率,并且使网络更不容易陷入过度兴奋状态。在当前刺激强度下,最高频率(25Hz,beta 节律)出现在 $V_{RR}=0.4$,$V_{RF}=1$ 处。

图 3.24(b)显示了同步性和连接强度之间的关系。可以看到在图 3.24(a)白色区域的对应处也是白色,说明持续兴奋下的功能柱同步性几乎为 0,除此以外,同步性系数 $\kappa$ 随着兴奋型连接强度的增加而增加,随着抑制性连接强度的增加而减少。同步性最强的区域总是出现在持续兴奋的边缘,此处对应图 3.23(c)中大振幅的振荡。




(a) 在不同连接强度下,功能柱活动的振荡频率可在3~20Hz之间变化($P_{stim}$=0.025)

(b) (a)中相应功能柱活动的同步性($P_{stim}$=0.025)

(c) 在不同 $V_{RR}$ 下振荡频率与输入强度的关系,提高刺激强度使功能柱的输出频率增大($V_{RF}$=0.4)

(d) (c)中对应的同步性变化,图中可以看出要使功能柱发生同步振荡,需满足有刺激的最小强度($V_{RF}$=0.4)

图 3.24 兴奋性和抑制性回路的耦合强度及输入强度对功能柱同步振荡频率和同步性的影响

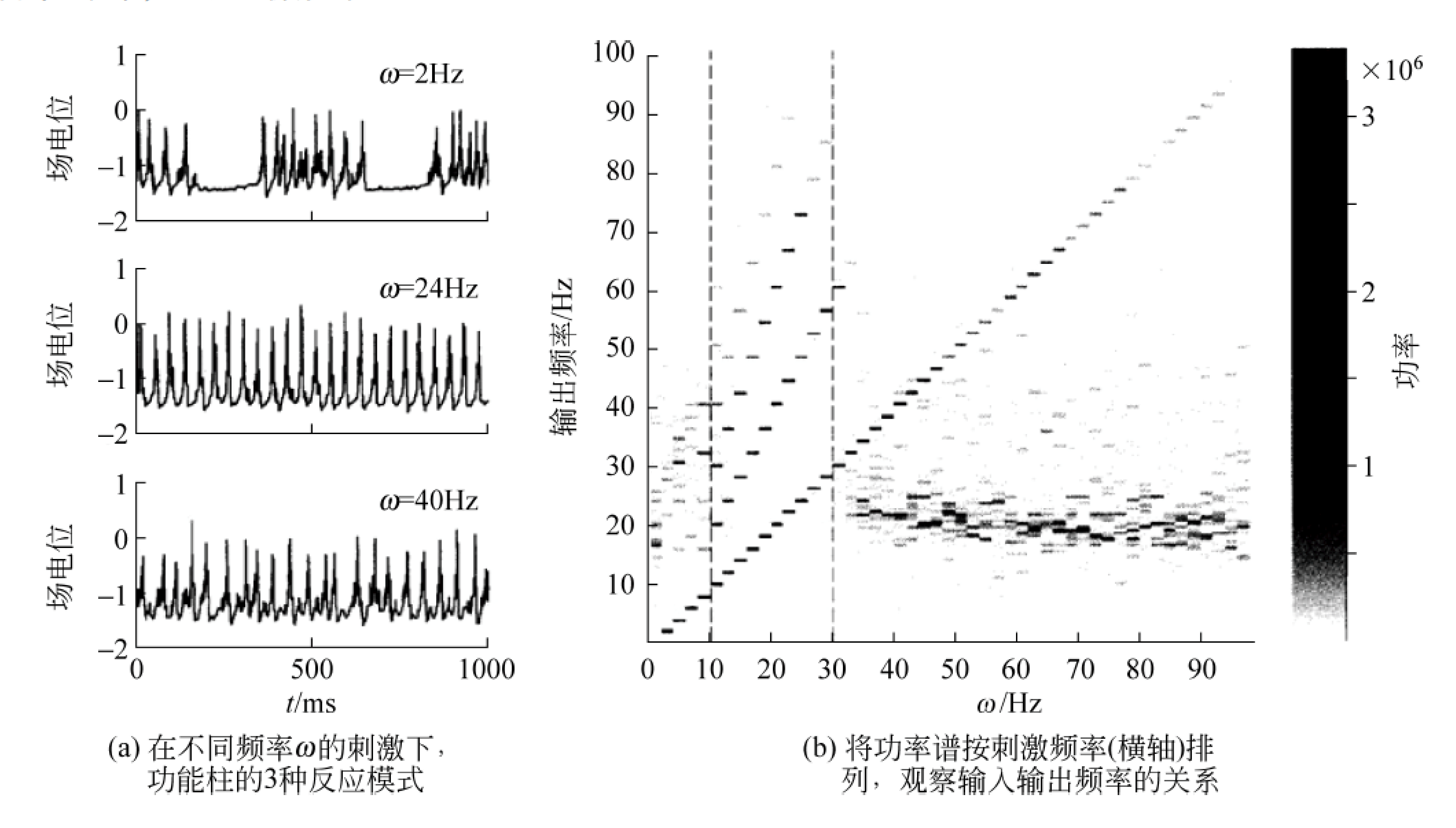
## 2. 同步性与刺激强度的关系

我们接着探讨网络行为与刺激强度的关系。刺激强度 $P_{stim}$ 从 0.01 到 0.05 逐渐变化,观察网络的周期性和同步性(见图 3.24(c)(d)),同时在不同 $V_{RR}$ 下刺激强度的作用。在图 3.24(c)和(d)当中都可以明显地看到,$P_{stim}$ 必须大于一个临界值约 0.016,才有同步振荡的出现,而且不同的 $V_{RR}$ 值对这个临界值几乎没有影响,说明使功能柱进入同步状态需要一个刺激的最小强度。当 $P_{stim}$ 超过临界值,则同步性迅速上升,很快达到最大值,$V_{RR}$ 越大,这个爬升的速度就越快,同步性 $\kappa$ 的峰值出现的就越早,当超过峰值以后,$P_{stim}$ 的增大均使 $\kappa$ 下

降，而且不同的 $V_{RR}$ 下曲线非常接近。说明要使网络较好地同步，需要最佳的 $P_{stim}$。$P_{stim}$ 过大，将使模型进入持续发放状态，同步性降为 0。网络振荡的频率随着 $P_{stim}$ 增大而单调递增，从 6Hz 变化到 43Hz，覆盖了 $\alpha$ 频率到 $\gamma$ 频率。

#### 3. 网络对周期性刺激的反应

给网络施加正弦刺激，以考察其行为对周期性输入的反应。$P_{stim}$ 由式(3.100)描述。首先我们选取了以下参数 $V_{RR}=0.3$，$V_{RF}=0.8$，$E_{stim}=0.025$，$A_{stim}=0.01$ 为不变量，当 $\omega=0$Hz 时，即如前面所讨论的恒定刺激，在恒定刺激下的频率称为此参数条件下的功能柱的固有频率。刺激的频率 $\omega$ 从 0 到 100Hz 间变化，功能柱的输出表现出 3 种反应模式(图 3.25(a))。在图 3.25(b)中输出的功率谱按 $\omega$ 排列，表示输出频率和输入频率之间的关系。可以看出当 $\omega<10$Hz 时，场电位的固有频率叠加在缓慢振荡上。在频谱中，在对角线上的峰表示输出频率中对应输入频率的成分，在 30Hz 处另有频率响应，也就是模型在最大输入 $P_{stim}=E_{stim}+A_{stim}=0.04$ 时的固有频率。当 $\omega>10$Hz 且 $<30$Hz 时，频谱中只有输入频率可见，网络的固有频率被外界输入同步化。当 $\omega$ 大于 30Hz，场电位中又可见到两个波叠加的情况。此时频谱中除输入频率外，另有 25Hz 左右的响应，并且随着 $\omega$ 增加，该峰逐渐向低频偏移并逼近 20Hz，也就是功能柱在平均输入 $P_{stim}$ 下的固有频率，同时频谱中输入频率的峰值也逐渐减小。



(a) 在不同频率ω的刺激下，功能柱的3种反应模式

(b) 将功率谱按刺激频率(横轴)排列，观察输入输出频率的关系

图 3.25　功能柱对周期性输入的反应

## 3.11　神经元集群的编码和解码

### 3.11.1　概述

脑的信息编码研究由来已久，在上世纪 20 年代阿德里安(Adrian)提出神经动作电位的

概念,他在 20 世纪 30 年代进行的实验工作,为揭示大脑信息处理提供了一些基本线索。从 1949 年赫布提出的经典细胞群假设[311],到 1972 年巴洛(Barlow)的单个神经元的编码假设[58],以及 1996 年藤井(Fujii)等人提出的动态神经元的集群时空编码假设[249],不同观点间的争论仍在继续。其中重要的一个问题是:是单个神经元还是神经元集群编码刺激信息是神经元动作电位出现的明确时间还是电位脉冲的平均发放速率携带信息。由于神经系统的高度复杂性,利用现有的实验手段还不能彻底解决神经元信息编码原理。但是现在已有越来越多的实验结果提示我们,神经系统中信息的编码与处理在很大程度上是在特定的发放频率与发放模式的框架下,通过大量神经元构成的集群编码活动完成的。在神经元集群中,每个神经元的活动特性都有其自身的特点,因而存在一定的差异性。然而,它们通过短暂的相关性活动与其他神经元进行相互协调,以神经元群体的整体活动或神经元活动的动态相关关系为特征,来实现对多种信息的并行处理和传递。

目前集群编码作为大脑信息处理的一种通用模型,主要是由于单个神经元对刺激的反应是充满噪声的并且对刺激值的变化缺乏灵敏性这样的实验事实,因此具有代表性的单个神经元所携带的信息是非常低的。大脑要克服这种局限,就必须将信息分配于拥有大量数目的神经元集群来共同携带关于刺激的精确信息[625]。集群编码的一个关键特性在于它的鲁棒性和可塑性,由于信息的编码在许多神经元共同活动的基础上得以完成,因此单个神经元的损伤不至于在太大程度上影响编码过程。集群编码还具有其他一些优点,例如可以降低噪声水平,并有助于短时程信息储存的形成等;同时这种编码方式也具有复杂性和非线性等特性。

神经元集群编码的一种方式是经典放电率模型意义下的群体编码。在早期的研究工作中,人们通过单位时间内动作电位的放电次数,对给定刺激作用下神经元的响应进行描述。这个测量值称为放电率,它一般由刺激诱导的放电率的平均响应,典型情况下呈钟形分布,和叠加于其上的噪声部分组成,噪声在每次测量时都有变化。早期人们的注意力主要集中在放电率上,因为该参量较为简单,易于测量且易于理解。虽然不能包含其所代表的各种各样的神经信息,比如刺激强度的大小;虽然仍没有完全了解神经信息是如何通过动作电位来编码的,但是动作电位作为神经信息编码的基本单位是无疑的。当然响应的其他方面的特性,譬如动作电位发生的精确时间关系,即放电序列模式对信息编码来说同样具有重要的意义。

考虑在不同噪声水平和神经元相关性的影响下,通过给定刺激条件下观察记录到的神经元活动,建立描述外界刺激与神经元响应间的对应关系的概率模型已成为研究集群编码的普遍方法,基于这种广泛的共识,产生了大量分析集群编码与解码的研究。有研究如何量化集群编码的信息,罗尔斯(E T Rolls)等人运用信息理论方法量化猕猴下颞叶可视皮层中神经元集群的编码信息,研究发现编码信息大致随着神经元数目线性增长[629];佛朗哥(L Franco)等人研究发现编码信息不仅存在于单个神经元的放电次数中,也受神经元活动的互相关性的影响,指出由独立于刺激的神经元互相关性影响产生的基于放电率的编码信息的冗余性是很小的,同时依赖于刺激的互相关性影响对下颞叶可视皮层中的编码信息的贡献也极小[231]。有从熵和互信息来探讨集群编码与解码的特性[548],康(K Kang)利用统计动力学方法计算神经元集群的互信息,结果发现在离散刺激下互信息随着集群规模指数饱和,而饱和指数恰恰是不同刺激诱导的神经元响应概率中的 Chernoff 距离[360];潘宁斯基

(L Paninski)探讨了集群编码中熵与互信息的非参数估计问题[548]。从计算角度来看，Fisher 信息量是衡量编码与解码精度的常用手段和有效指标[594]。例如阿博特(L F Abbott)和戴安(P Dayan)研究了互相关神经元的放电率差异性对集群编码精度的影响，通过对编码过程的 Fisher 信息量的计算，指出一般情况下只要集群的规模足够大的话，相关性并不限制编码精度的增长，并且讨论了某些情形下(例如加性噪声和乘性噪声)相关性怎样提高编码精度[5]；萨默波林斯基(Haim Sompolinsky)等通过 Fisher 信息量的计算发现，相对于不相关的神经元集群，正相关性降低神经网络的估计能力，信息容量随集群中神经元数目增长而趋于饱和，而负相关性则大大提高集群的信息容量[733]；托约竹米(T Toyoizumi)研究了对基于峰电位的集群解码(不考虑峰电位发生时刻)中损失的 Fisher 信息量的估计问题[760]。

贝叶斯推理法则是研究神经元集群编码与解码的关键，是量化编码与解码行为的重要方法。早在 1998 年泽梅尔(R S Zemel)就给出了贝叶斯原理框架下神经元集群编码与解码活动的概率解释[831]，比较了在外界刺激诱导条件下神经元放电活动的泊松模型、KDE (Kernel Density Estimation)模型与扩展泊松(Extended Poisson)模型的性能，包括编码、解码、似然度与误差分析比较。近年来的理论研究表明大脑中包括编码与解码的神经计算过程类似于贝叶斯推理过程[595]；目前贝叶斯方法已被成功用于感知与感觉控制的神经计算理论，并且心理物理学上不断涌现的证据也表明大脑的感知计算是贝叶斯最优的，这也导致了尼尔(D C Knill)等人称之为贝叶斯编码假说[374]。从记录到的神经元放电活动中重构外界刺激或刺激的某些特性，贝叶斯推理为揭示这样的解码过程行为提供了可能[182,183]。葛杨(Yang Ge)和蒋文新(Wenxin Jiang)探讨了采用逻辑回归混合模型的贝叶斯推断的一致性[255]。

神经元集群编码与解码是神经信息处理的关键问题，是揭示大脑工作机理的理论框架，它的发展能够促进人们对脑的总体功能的认识，为研究更为复杂的高级认知功能提供基本理论与指导方法。基于贝叶斯原理的编码与解码方法能够从总体上大致揭示神经系统信息处理过程的特性，为脑的工作机理作出客观合理的数学解释。

## 3.11.2 熵编码理论

信息熵是随机事件的不确定性及信息量的量度，又可以反映由偏度、峰度代表的不规律性，因此适合表达非线性的神经元发放编码。熵编码方法包括：二值串方法、直接方法、渐近线无偏估计方法、神经元发放间隔方法、神经元发放对数间隔和互信息方法等。

### 1. 二值串方法

1952 年麦凯(Mackay)和麦克罗奇(McCulloch)首次把熵的概念用于神经编码的表达，提出了二值串熵编码，该方法的原理是将在 $\Delta t$ 时间内包含 $r$ 个发放的发放序列离散成多个带宽为 $\Delta T$ 的窄带，在每个窄带中根据有无发放分别用“1”和“0”表示编码。这种“1”和“0”序列的熵表达式为：

$$S[c] = \log_2({}^{n}C_r) \tag{3.102}$$

式中

$${}^{n}C_r = \frac{n!}{r!(n-r)!} \tag{3.103}$$

$n$ 为二值串的长度,即窄带 $\Delta T$ 的个数;$r$ 为二值串中“1”的个数。

一般二值串都很长,难于计算其长度,所以常用近似公式。当 $n$、$r$、$(n-r)$ 都很大,且 $r \ll n$ 时式(3.102)可以转化为:

$$S[c \mid M, \Delta\tau] \approx \log_2 \frac{e}{M\Delta\tau} \tag{3.104}$$

式(3.104)中,$M$ 为平均发放频率;$\Delta T$ 为窄带的带宽;e 为自然对数。

二值串方法可以近似计算神经元发放序列的熵,但该方法存在“带宽问题”的局限性,即对于不同的发放序列,离散的带宽 $\Delta\tau$ 并没有明确的标准来确定,因此导致计算的熵值可能与实际不符[621]。

### 2. 直接熵编码

1997 年普林斯顿大学的史帝文柯(Steveninck)等人提出了直接熵编码方法[779]。该编码方法的思路是基于在一个刺激前后的特定时间段内发放的时间之间的联系。编码表达的形式为:将发放序列离散成多个带宽为 $\Delta t$ 的窄带,在每个窄带中根据有无发放分别用“1”和“0”表示编码,在某一个设定的常数 $L$ 下,由 $L$ 个相邻窄带编码的不同组合表达不同的“词”,再根据不同的“词”在整个发放序列中的概率分布计算出发放序列的熵

$$S = -\frac{1}{L\Delta\tau}\sum_{i=1}^{n} p(w_i)\log_2 p(w_i) \tag{3.105}$$

式(3.105)中,$L$ 为设定的常数;$n$ 为发放序列中包含的“词”的个数;$w$ 表示长度为 $L$ 的第 $i$ 个“词”;$p(w_i)$为第 $i$ 个“词”在发放序列中出现的概率。

直接方法的优点是熵与窄带分布和刺激特性无关,但是也同样有如在二值串方法中存在的“带宽问题”;同时还存在“数据采样问题”,即为了准确估计“词”出现的概率,需要在实验中采集大量的数据,这常常难以实现。

史帝文柯等人于 1997 年对苍蝇视觉中枢的运动神经元施加恒定刺激并记录神经元的响应,再应用直接方法用熵定量计算了神经元发放序列的可重复性和变异性,从而研究了在某类特定刺激下神经信息编码的方式。

### 3. 渐近线无偏估计熵编码

2002 年维克特(J Victor)提出了熵编码的渐近线无偏估计方法。方法的基本原理是将神经元发放序列“嵌入”一个矢量空间集,再从欧几里得空间中的有限数据集中估计出此发放序列连续分布的熵[786]。

具体表达编码的方法是:将神经元发放序列携带的信息分成两个分布:一个是关于发放计数的分布,即在 $\Delta t$ 时间内发放个数的分布;另一个是关于发放发生时间的分布。因此可以用离散为窄带的方法来表达发放计数携带的信息编码,用非窄带方法估计发放时间携带的信息编码。该方法使非线性发放序列线性、连续地嵌入矢量空间集,避免了“带宽问题”,但是目前该方法的编码结果和实际生理机制相差甚远,与实际应用尚有距离。

### 4. 神经元发放间隔的熵编码

神经元发放间隔熵编码的主要思路是应用神经元发放脉冲间隔(Inter-Spike Interval,

ISI)来表达编码。首先用发放间隔图(Inter-Spike Interval Histogram,ISIH)来表示 ISI 的分布情况,再找出合适的间隔概率密度函数 $f_w(w|I)$来拟合 ISIH,再由 $f_w(w|I)$计算出 ISI 的熵:

$$S[w]=-\int_{S_w} f_w(w \mid I)\log_e f_w(w \mid I)\mathrm{d}w \tag{3.106}$$

式(3.106)中,$w$ 表示 ISI 序列;$f_w(w|I)$为一个合适的条件概率密度函数,例如可以选择 gamma 函数、正态函数或高斯概率密度函数等[468];$S_w$ 是 $f_w(w|I)$的定义域。

ISI 熵编码方法可以识别出被"直接方法"所忽略的信息。但是,因为 ISI 的均值就是发放频率的倒数,所以 ISI 熵编码方法在本质上还是一种基于频率的编码方法。目前,应用 ISI 的熵来表达编码已应用于很多神经元和细胞发放序列的分析中,例如 1996 年马修(P. Matthews)用 ISI 熵编码方法分析了人类多种不同肌肉的运动神经元的发放编码,用以解释突触噪声对刺激响应的影响;2001 年凌(G Leng)等将它用于解释神经元对持续增强的突触输入的抑制效应[413];2004 年魏特摩(D Wetmore)和贝克(S Baker)对猴运动皮质细胞发放编码应用了 ISI 的熵编码方法,得出了猴运动皮质的单个神经元在发放动作电位 30ms 后再次发放的概率逐渐增加的结论[808]等等。该方法存在的问题是,选择 ISIH 的拟合函数比较困难,拟合精度有待提高。

### 5. 神经元发放对数间隔的熵编码

神经元发放对数间隔的熵编码在 ISI 熵编码的基础上,用 ISI 的对数值代替 ISI 值作为 ISIH 的横坐标,构造出对数发放间隔图,再用合适的间隔概率密度函数 $f_w(w|I)$来拟合对数发放间隔图。布姆勃拉(G Bhumbra)等人提出用双峰高斯函数很好地拟合了多种神经元发放序列的对数间隔[75]。

如果 $x$ 的单位是 $\log_e$(ms),则双峰函数 $f(x)$可以用加权系数 $c$,两个均值 $\mu_1$ 和 $\mu_2$,两个标准差 $\sigma_1$ 和 $\sigma_2$,包含以上 5 个参数的拟合函数如下:

$$f(x)=\frac{c}{\sigma_1\sqrt{2\pi}}\mathrm{e}^{-(x-\mu_1)^{2/2\sigma_1^2}}+\frac{1-c}{\sigma_2\sqrt{2\pi}}\mathrm{e}^{-(x-\mu_2)^{2/2\sigma_2^2}} \tag{3.107}$$

应用 Levenberg-Marquardt 迭代算法来优化以上 5 个参数,再用 K-S 检验参数 $D$ 来衡量数据和双峰函数的符合程度。式中 $D$ 代表的是期望累积密度和实际累积概率之间的最大差距,期望累积密度函数 $f(x_i \leqslant x)$的表达式为:

$$f(x_i \leqslant x)=\int_{-\infty}^{x_i} f(x)\mathrm{d}x \tag{3.108}$$

再将 $f(x)$代入式(3.106)计算出对数间隔熵。

对数间隔图的局限性是,它不能表达间隔发生次序的信息,因为这些信息可能是很重要的,所以进一步用相邻间隔的互信息来衡量一个发放前后两个间隔之间的信息。计算互信息的过程如下。

(1) 用对数间隔 $Y$ 及其前驱 $x$ 画出对数间隔散点图。

(2) 用一个标准差为对数间隔标准差六分之一的二维高斯核与散点图进行卷积来平滑数据。

(3) 构造联合对数间隔图。

(4) 得出相邻间隔的联合概率质量分布 $P(x_i, y_i)$。

(5) 由 $P(x_i, y_i)$计算出前驱间隔 $P(x_i)$及后驱间隔 $P(y_i)$的边缘概率分布：

$$P(x_i) = \sum_{j=1}^{N_y} P(x_i, y_i) \tag{3.109}$$

$$P(y_i) = \sum_{i=1}^{N_x} P(x_i, y_i) \tag{3.110}$$

式中 $N_y$ 和 $N_x$ 分别是 $P(y_i)$和 $P(x_i)$的窄带数。$P(x_i)$的熵 $S(X)$可表示为：

$$S(X) = -\sum_{i=1}^{N_y} P(x_i) \log_2 P(x_i) \tag{3.111}$$

(6) 联合概率质量分布 $P(x_i, y_i)$的联合熵表示为：

$$S(X,Y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} P(x_i, y_i) \log_2 P(x_i, y_i) \tag{3.112}$$

相对熵 $D(X,Y \| XY)$表示为：

$$D(X,Y \| XY) = S(X) + S(Y) - S(X,Y) \tag{3.113}$$

相对熵 $D(X,Y \| XY)$是相邻间隔互信息的近似值。但是此近似值很可能偏高，所以需要引入随机化方法，通过随机和重列进行校正。并可以用 Monte-Carlo 法验证互信息是不是显著的大于零。

2001 年菲尔霍(D Fairhall)将发放对数间隔分析方法应用于苍蝇视觉系统对快速变化的刺激的响应分析中，揭示了短发放间隔在神经信息表达中的重要性[216]。2004 年布姆勃拉应用发放对数间隔分析方法对鼠视上核神经元核周细胞、抗利尿激素细胞和催产素细胞的发放模式进行了定量比较[75]。该方法也存在对数发放间隔图很难精确拟合的局限性。

上述熵编码方法，二值串熵编码存在对于不同的发放序列，离散的带宽 $\Delta f$ 没有明确标准，因此导致计算的熵值可能与实际不符的局限性；直接熵编码存在为了准确估计“词”出现的概率，需要在实验中大量采集数据，又常常难以实现的局限性；渐近线无偏估计熵编码存在与实际应用距离较远的局限性。与以上四种方法相比，应用神经元放电序列对数间隔熵和互信息熵编码，通过熵值与互信息的结合，可以更好地表达神经元放电的编码。熵编码研究的发展动态之一是结合对数间隔熵和互信息熵，更好地揭示神经元放电编码的机制。

### 3.11.3 贝叶斯集群编码

在给定的随机模型下，编码与解码通过贝叶斯法则相互联系。这里 $r$ 表征单个神经元或集群神经元在刺激条件下的响应，刺激用相位参量 $\theta$ 表示。$n$ 个神经元的响应记为 $r=(r_1, r_2, \cdots, r_i)$，这里 $r_i, i=1,2,\cdots,N$ 表示用峰电位计数的第 $i$ 个神经元的放电率，当然除了放电率，还有其他描述神经元响应的参量。下面我们引入描述 $N$ 神经元活动的概率函数。

$P(\theta)$：用参量 $\theta$ 表示刺激的概率，它经常被称为先验概率或先验知识。

$P(r)$：实验中记录到的响应 $r$ 的概率。

$P(r,\theta)$：刺激 $\theta$ 与记录到的响应 $r$ 的联合分布概率。

$P(r|\theta)$：由刺激 $\theta$ 激发响应 $r$ 的条件概率。

$P(\theta|r)$：在响应 $r$ 被记录的条件下刺激为 $\theta$ 的条件概率。

这里需要注意的是，$P(r|\theta)$是在刺激取值 $\theta$ 时观察到放电率为 $r$ 的概率，而 $P(r)$表示放电率为 $r$ 时的概率，它不依赖于刺激取的某个特定值，由此 $P(r)$可以用 $P(r|\theta)$在所有刺激值的概率的权重之和表示，即

$$P(r) = \sum_{\theta} P(r \mid \theta) P(\theta) \tag{3.114}$$

同理有

$$P(\theta) = \sum_{r} P(\theta \mid r) P(r) \tag{3.115}$$

由条件概率的定义我们知道关于刺激 $\theta$ 与响应 $r$ 的联合概率表示为

$$P(r,\theta) = P(r \mid \theta) P(\theta) = P(\theta \mid r) P(r) \tag{3.116}$$

由此有

$$P(\theta \mid r) = \frac{P(r \mid \theta) P(\theta)}{P(r)} \tag{3.117}$$

这里假定 $P(r) \neq 0$，此即为从 $P(\theta|r)$到 $P(r|\theta)$的贝叶斯推理理论。编码可以通过一组关于所有刺激与响应的概率 $P(r|\theta)$描述，另一方面，解码某个响应，相当于获取概率 $P(\theta|r)$。由贝叶斯理论，$P(\theta|r)$可由 $P(r|\theta)$获取，但却需要刺激概率 $P(\theta)$。因此解码需要用到实验中或自然发生的刺激的统计性质的知识。

### 3.11.4 贝叶斯集群解码

贝叶斯方法在解码问题上的好处在于促使我们作出明确的假设，在处理似然度时我们可通过先引入一些简单的假设，然后研究怎样使假设越来越实际。这里先给出刺激 $\theta$ 下 $n$ 个神经元放电率分别为 $r_i$ 的联合似然度为 $P(r_1,r_2,\cdots,r_n|\theta)$。如果我们知道刺激 $\theta$，并且单个神经元的放电率依赖于 $\theta$ 且独立于其他神经元的放电率，则假定不同神经元放电率在刺激 $\theta$ 下是条件独立的，则有

$$P(r_1,r_2,\cdots,r_n \mid \theta) = \prod_{i=1}^{n} P(r_i \mid \theta) \tag{3.118}$$

这个假定意味着所有神经元放电率的联合似然度就等于各自似然度的乘积。

从数学意义上讲，编码过程可用在给定刺激 $\theta$ 下的条件概率 $P(r|\theta)$描述，而解码就是从观察到的集群活动 $r$ 中推断刺激 $\theta$ 的值。文献[835]中指出大多数的解码方法可以用极大似然推断 MLI(Maximum Likelihood Inference)或最大化后验估计 MAP(Maximize a Posteriori)来系统的表述，具体通过选择适当的似然函数和关于刺激的先验分布来实现。它们可以归纳为如下的一致方法。

由贝叶斯法则，在给定 $r$ 的条件下关于刺激 $\theta$ 的后验分布 $P(\theta|r)$为

$$P(\theta \mid r) = \frac{P(r \mid \theta) P(\theta)}{P(r)} \tag{3.119}$$

这里 $P(r|\theta)$为似然函数，$P(\theta)$为关于 $\theta$ 的分布，表征先验知识，$P(r)$为标准化因子。MAP 就是通过对后验分布的对数 $\ln P(r|\theta)$的最大化来实现对刺激的估计，即

$$\hat{\theta} = \mathrm{argmax}_{\theta} \ln P(\theta \mid r) = \mathrm{argmax}_{\theta} [\ln P(r \mid \theta) + \ln P(\theta)] \tag{3.120}$$

当先验知识 $P(\theta)$未知或平直的，MAP 还原为 MLI

$$\widehat{\theta} = \mathrm{argmax}_{\theta} \ln P(r \mid \theta) \tag{3.121}$$

注意在解码阶段,由估计者假定的 $P(r|\theta)$要是等于实际的模型 $Q(r|\theta)$,我们就称该估计者运用了一个可置信模型;当 $P(r|\theta)$不等于 $Q(r|\theta)$,则称为非置信模型。运用非置信模型主要基于两个原因,一是解码系统经常不知道编码系统的精确信息,因此不得不使用非置信模型,特别是对实验中的数据分析;二是通过合适的简化的非置信模型,可以在不牺牲太多的解码精度的要求下,实现对计算成本的大幅降低。

极大似然推理类型估计量 $\widehat{\theta}$ 的获得可通过最大化似然对数 $\ln P(r|\theta)$得到,即解方程 $\nabla \ln P(r|\widehat{\theta})=0$,这里$\nabla P(s)$表示 $\mathrm{d}P(s)/\mathrm{d}s$,$P(r|\theta)$称为解码模型。这里考虑三种解码模型,它们都采用极大似然推理,定义如下。

(1) 置信模型 $P_F(r|\theta)$(Faithful Model),它用到所有的编码信息,解码模型就是真实的编码模型,即 $P_F(r|\theta)=Q(r|\theta)$,这种方法简记为 FMLI。

(2) 非置信模型 $P_U(r|\theta)$(Unfaithful Model),它运用描述神经元响应活动的调置函数的信息,但是却忽略神经元之间的相关作用,简记为 UMLI。

(3) 矩法(也称为矢量法)模型 $P_C(r|\theta)$,它并不涉及运用编码过程的任何信息,但却对调置函数作出了粗糙的简单假定,同时忽略神经元之间的相关作用,简记为 COM。

这里需要指出的是利用非置信模型(例如 $P_U(r|\theta)$或 $P_C(r|\theta)$)有重要的意义。当实验研究人员从记录到的数据中重建刺激时,它们实际上运用的是非置信模型,因为真实的编码过程是无从知道的。进一步讲,真实的神经元之间的相关性经常是复杂的并且随着时间变化,导致大脑很难储存并利用所有这些信息,因此基于非置信模型的极大似然推理(忽略信息的某些方面)是解决这样的信息“灾难”的关键。

# 第 4 章<<

# 心智模型

心智(Mind)是人类全部精神活动,包括情感、意志、感觉、知觉、表象、学习、记忆、思维、直觉等,人们用现代科学方法来研究人类非理性心理与理性认知融合运作的形式、过程及规律。建立心智模型的技术常称为心智建模,目的是探索和研究人的思维机制,特别是人的信息处理机制,同时也为设计相应的人工智能系统提供新的体系结构和技术。

## 4.1 心智建模

智能科学的类脑智能研究与心智的计算理论有密切的关系,一般采用模型表示心智是怎样工作的,理解心智的工作机理。1980 年,纽厄尔首先提出了心智建模的标准[529]。1990 年,纽厄尔把人类心智描述为一组功能约束,提出心智的 13 条标准[531]。2003 年,在纽厄尔 13 条标准的基础上,安德森等提出 Newell 测试[32],判断人类心智模型要满足的标准以及更好工作所需要的条件。2013 年,文献[691]分析了心智模型的标准。为了更好地构建心智模型,这里综合讨论心智建模的标准。

### 1. 灵活的行为

纽厄尔在 1990 年的《统一的认知理论》一书中[531],重申第一条标准“行为灵活地作为环境的函数”。在 1980 年纽厄尔认为,心智的普适性是最重要的标准。对纽厄尔来说,人类行为的灵活性反映心智的普适性;行为是环境的任意函数。

纽厄尔认为人类心智的灵活性与现代计算机的特点相同。他认识到这种能力的创建与普适计算的形式化表示难度相同。例如,记忆有限性使人不能等效于图灵机(具有无限的带子),并且频繁的移动影响人展示完美的行为。不过,纽厄尔认识到人类认知的真正的灵活性应该与计算的普适性相同,就像现在把现代计算机描述为等效的图灵机一样,尽管它有物理限制和偶然的误差。

当计算的普适性是人类认知的事实时,即使计算机具有专业化的处理器,它不应该看作仅执行各种各样具体的认知功能。而且,表明人们学习一些事情比其他设备容易得多。在语言领域里强调“自然语言”,自然语言的学习比非自然语言容易得多。常用的人工制品只是非自然系统的一小部分。当人们可能逼近计算普适性,只是获得可计算函数的极小部分,并且给予执行。

评价分级:一种理论是否好,它应该相对直截了当地回答它是否是通用可计算的。这并不是说,该理论会容易发现一切等同的事情,或者说人的表现将永远不会出现错误。

### 2. 实时性

对于一个认知理论来说,仅有灵活性还是不够的,它必须解释人怎样能在实时的条件下完成,这里"实时"表示人执行的时间。作为对神经网络了解的人,认知增加了限制。实时是对学习和执行同样的限制。如果学习某些事情,需要花费毕生的时间,原则上这种学习是不好的。

评价分级:如果一个理论是特定的约束,可以快速地实现处理的过程,那么对于任何人类认知的特定情况下,它是否能够决定实时实现是不重要的。这不可能证明该理论满足所有人类认知情况的实时约束,能否满足必须看具体情况。

### 3. 自适应的行为

人类不只是进行奇妙的智力计算,而且要选择满足他们需要的计算。1991年,安德森提出有两级自适应性:一级是系统结构的基本过程和相关的形式,提供有用的功能;另一级是整个系统看作一个整体,它的整个计算是否满足人们的需要。

评价分级:纽厄尔认为短时记忆模型不具备自适应的功能。在SOAR系统中,可以探讨它的机制是否允许行为具有真实世界的功能。

### 4. 大规模的知识库

人的自适应性关键的一点是可以访问大量的知识。也许人的认知与各种各样的"专家系统"的最大区别在于,大多数情况下人有采取适当行动的必要知识。不过,大规模的知识库会产生问题。不是所有知识都同样可靠或者同等相关。对于当前情况来说,相关的知识能迅速变得不相关。成功地储存全部知识和在合理时间内检索相关的知识可能有严重问题。

评价分级:为了评估这一标准,需要确定性能随知识规模变化的情况。同样,理论是否被良好规定,这个标准遵循实际的分析。当然系统的规模对性能是有影响的。

### 5. 动态行为

在现实世界中,不像求解迷宫、河内塔问题那么简单。世界的变化不是我们能期望的,也不能控制。即使人们的行为想控制世界,但会有想不到的效果。处理动态的、无法预测的环境是全部生物体生存的前提。鉴于人们已经为他们自己的环境建立复杂性分析,对动态反应的需要主要是面临认知问题。处理动态行为需要有感知和行动的理论以及认知理论。情景认知的工作强调由于外部世界的结构认知怎样出现[285,742]。这一立场的支持者争辩说,全部认知是外部世界的反映。这与较早认为认知可以无视外部世界是一个鲜名的对照[149,150]。

评价分级:怎样才能创建一个系统测试"动态意外"?当然,典型的实验室这种乏味的工作是通过试验进行。适当的测试需要将这些不受控制的环境插入到系统中。在这方面,一个很有前途的类似测试是在这些系统中构建认知智能体,并且插入到真实的或合成的环境里。例如,纽厄尔的SOAR系统在空军任务的模拟中成功地模拟飞行员,涉及包括飞行员在内的5000个智能体[350]。

### 6. 知识集成

纽厄尔称这条标准为“符号和提取”。纽厄尔对这条标准的评论出现在他的《统一的认知理论》一书中：“心智能够使用符号和提取。我们知道，只是观察自己。”他好像从未承认，这个问题有什么争论。纽厄尔认为符号、方程式之类的外部符号，其存在大概很少有争论。他认为符号就是表处理语言的具体事例。许多符号没有直接的意思，与哲学的讨论或者计算的效果不同。在纽厄尔意义上，符号作为分级标准，不可能安装。但是，如果我们注意他的物理符号的定义，就会明白这条标准的合理性[531]。

评价分级：建议评审该理论是否能够产生人们智能活动的能力特点组合，比如推理、归纳、隐喻和类比。操纵系统产生任何特定的推论始终是可能的，是有限制的正常能力的智力结合。系统应该能够重现日常人们显示的智力组合，这源于动作始终发自身体的局部，在有限空间内有限的知识是可编码的，人类心智包含大量的知识。因此，编码知识必须在空间上传播出去，它被存储在处理需要它的地方。符号是实现远程访问所需的工具[531]。符号把人类理性概念有关的最密切知识联系在一起进行推理。

### 7. 自然语言

自然语言是人类符号操纵的基础。相反，在何种程度上符号操纵是自然语言的基础。纽厄尔认为语言取决于符号操纵。

评价分级：作为测试的一部分，并且是重要的部分，社会上已建立语言处理的测试，有点像读一条消息，并且回答问题。这将涉及语法分析、理解、推理，以及对于有关当前文本的过去知识。

### 8. 意识

纽厄尔承认意识对整个人类认知的重要性。纽厄尔让我们考虑全部准则，而不是仔细挑选其中之一来考虑。

评价分级：科恩和斯科勒(J W Schooler)主编了《意识的科学方法》一书[151]，其中包含了潜意识感知、隐学习和记忆、元认知过程。建议一个理论对这一标准的衡量方法是其在产生这些现象的能力，解释为什么它们是人类的认知功能。

### 9. 学习

学习好像是人类认知理论的另一个不可控制的标准。一个令人满意的认知理论必须解释人类获得他们竞争力的能力。

评价分级：简单通过询问是否该理论能够学习似乎不够，因为人们必须能够具有许多不同种类的学习能力。建议采取斯奎尔(L R Squire)的分类[737]，作为测量理论是否可以解释人类的学习的范围。斯奎尔分类中的主要目录是语义记忆、情景记忆、技巧、启动和条件。他们可能并不明显地区分理论类别，可能有更多种的学习，但是这也代表了大多数人类学习的范围。

### 10. 发育

发育是纽厄尔最初列出的认知系统结构 3 条约束的第一条。尽管在假想世界里人们想

像相关联的功能与新的认知理论完全成熟,在现实世界中人类的认知被约束在有机体成长和响应体验中展开。

评价分级:与语言标准一样,发育分级也有问题,似乎人类的发育没有很好的整个维度表征。与语言相比,人类发育不是一种能力,而是一种约束,没有公认的测试标准,尽管世界上有很多对儿童发育测试的方法。

### 11. 进化

人类的认知能力,必须通过进化提升。已经提出各种内容-特定的能力,诸如以检测作弊者的能力[161],或在自然语言的约束[577],在人类进化史上的特定时间里发生进化。进化约束的变化是比较约束。人类认知的体系结构与其他哺乳动物有何不同?我们已经将认知可塑性作为人类认知的特征之一,语言也是确定的特征。什么是人类认知系统的基础和独特的认知属性?

评价分级:纽厄尔对采用哪些进化约束表示有些为难。进化约束分级是一个很深的问题,因为关于人类认知发展的数据很少。从对比人类的认知是怎样适应环境(标准3)来看,重建选择性能力的历史可以成为构建标准[227]。最好提出松散的理论,采用进化和比较的方法。

### 12. 脑

最后的约束是认知的神经实现。最近的研究进展,使可以用作研究认知约束理论的特定脑区的功能数据极大地增加。

评价分级:建立这种理论需要枚举和证明。枚举是把认知体系结构的构成模块映射到脑结构;证明是计算脑结构匹配认知体系结构的模块。令人遗憾,脑功能的知识还没有发展到这一步。但是,有足够的知识来部分地实现这样的测试。作为部分的测试,这种测试是相当苛刻的。

心智问题是一个非常复杂的非线性问题,我们必须借助现代科学的方法来研究心智世界。智能科学研究的是心理或心智过程,但它不是传统的心理科学,它必须寻找神经生物学和脑科学的证据,以便为心智问题提供确定性基础。心智世界与现代逻辑学和数学所描述的可能世界也有明显的区别:逻辑学和数学所描述的可能世界是一个无矛盾的世界,而心智世界则处处充满了矛盾;逻辑和数学对可能世界的认识和把握只能用演绎推理和分析方法,而人的心智对世界的把握则有演绎、归纳、类比、分析、综合、抽象、概括、联想和直觉等多种手段。所以心智世界比数学和逻辑学所描述的可能世界要复杂广大得多。那么,我们应该如何从有穷的、无矛盾的、使用演绎法的、相对简单的可能世界进入无穷的、有矛盾的、使用多种逻辑和认知方法的、更为复杂的心智世界呢?这是智能研究要探索的基本问题之一。

总之,智能科学是在当代科学技术发展成就的基础上,为提高人类认知水平,特别是提高人工智能水平而发展起来的一门新兴科学。智能科学的目标,就是要揭开人类心智的奥秘,它的研究不仅能够促进人工智能的发展,揭示生命的本质和意义,在促进现代科学特别是心理学、生理学、语言学、逻辑学、认知科学、脑科学、数学、计算机科学甚至哲学等众多学科的发展上,都有非同寻常的意义。

## 4.2　图灵机

英国科学家图灵(A M Turing)于 1936 年发表著名的《论应用于解决问题的可计算数字》一文[771]。文中提出思考原理计算机——图灵机的概念,推进了计算机理论的发展。1945 年图灵到英国国家物理研究所工作,并开始设计自动计算机。1950 年,图灵发表题为《计算机能思考吗?》的论文,设计了著名的图灵测验,通过问答来测试计算机是否具有同人类相等的智力。

图灵提出了一种抽象计算模型,用来精确定义可计算函数。图灵机由一个控制器、一条可无限伸延的带子和一个在带子上左右移动的读写头组成。这个在概念上如此简单的机器,理论上却可以计算任何直观可计算的函数。图灵机作为计算机的理论模型,在有关计算机和计算复杂性的研究方面得到广泛应用。

计算机是人类制造出来的信息加工工具。如果说人类制造的其他工具是人类双手的延伸,那么计算机作为代替人脑进行信息加工的工具,则可以说是人类大脑的延伸。

图灵机是一种无限记忆自动机,如图 4.1 所示。它由以下几个部分组成。

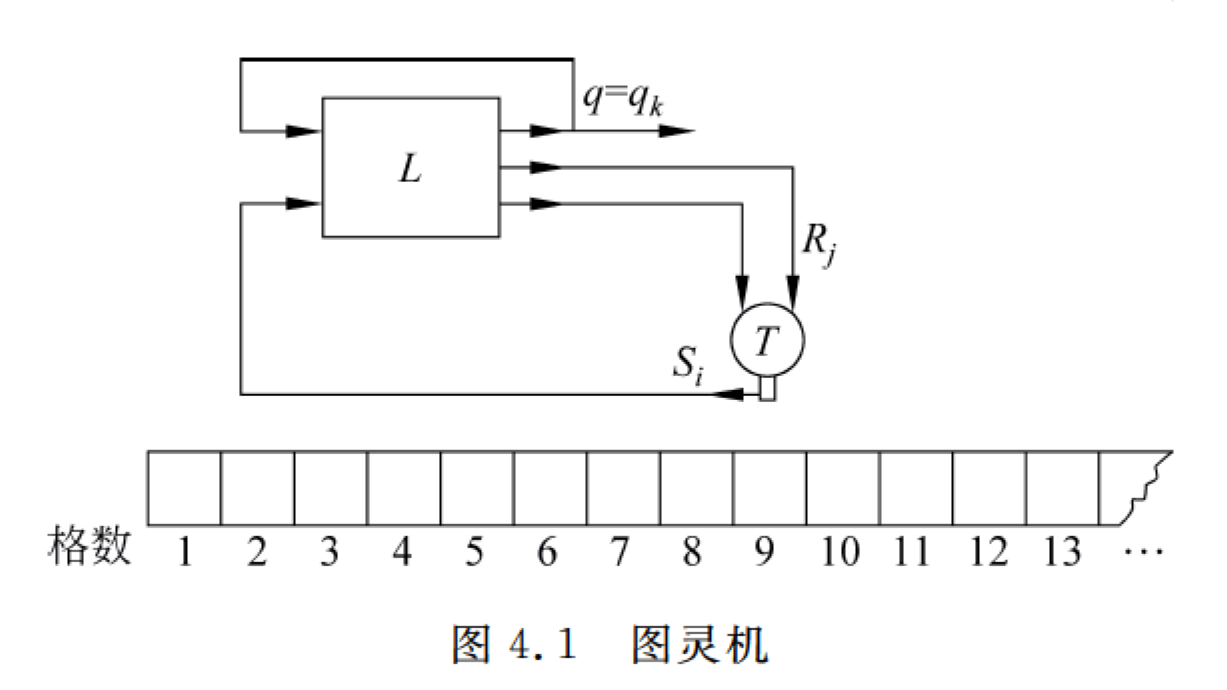



图 4.1　图灵机

(1) 一条无限长的纸带。纸带被划分为一个接一个的小格子,每个格子上包含一个来自有限字母表的符号,字母表中有一个特殊的符号表示空白。纸带上的格子从左到右依此被编号为 0, 1, 2,…,纸带的右端可以无限伸展。

(2) 一个读写头。该读写头可以在纸带上左右移动,它能读出当前所指的格子上的符号,并能改变当前格子上的符号。

(3) 一个状态寄存器。它用来保存图灵机当前所处的状态。图灵机的所有可能状态的数目是有限的,并且有一个特殊的状态,称为停机状态。

(4) 一套控制规则。它根据当前机器所处的状态以及当前读写头所指的格子上的符号来确定读写头下一步的动作,并改变状态寄存器的值,令机器进入一个新的状态。

纸带上的格子可以记录“0”或“1”。在带子上方移动一个读写磁头,它是由有限记忆自动机 $L$ 来控制的。自动机 $L$ 按周期工作,关于符号(0 或 1)的信息,由磁头从带子上读出,而馈给 $L$ 的输入。磁头根据在每个周期中从自动机 $L$ 得到的指令而工作,它可以停留不动或向左、向右移动一小格。与此同时,磁头从自动机 $L$ 接收指令,执行收到的指令,它就可以更换记录在磁头下面方格中的符号。

图灵机的工作唯一地决定于带子方格的初始存储和控制自动机的变换算子,这个算子

可以表示为转移表的形式。我们用 $S_i$($S_0=0$,$S_1=1$)表示磁头读出的符号;用 $R_j$[$R_0$(停止),$R_1$(左移),$R_2$(右移)]表示移动磁头的指令;用 $q_k$($k=1,2,\cdots,n$)表示控制自动机的状态,则表 4.1 给出了图灵机状态转移表。

**表 4.1 图灵机状态转移表**

| 输　入 | 状　态 | |
|---|---|---|
| | $S_0=0$ | $S_1=1$ |
| $q_1$ | $S_0,R_2,q_k$ | $S_1,R_1,q_m$ |
| $q_2$ | $S_1,R_0,q_s$ | $S_0,R_2,q_1$ |
| $q_3$ | $S_1,R_1,q_p$ | $S_0,R_2,q_2$ |

从表中看出,自动机 $L$ 的动作依赖于输入 $q$ 和它的状态 $S$。对于给定值 $q$ 和 $S$,将有 $q$,$R$,$S$,这三个量的某一组值与之对应。这三个量分别指明,磁头应在磁带上记录什么符号 $q$,移动磁头的指令 $R$ 是什么,自动机 $L$ 将变到什么新状态 $S$。在自动机 $L$ 的状态 $S$ 中至少应当有这样一个状态 $S^*$,对于这个状态来说,磁头不改变符号 $q$,指令 $R=R_0$(停止),而自动机 $L$ 仍处于停止位置 $S^*$。

图灵机的结构虽比较简单,但在理论上它却能够模拟现代数字计算机的一切运算,实现任何算法,因此可以看作是现代数字计算机的一种数学模型,可以通过对这种模型的研究揭示数字计算机的性质。

## 4.3 物理符号系统

我们把人看成一个信息加工系统,常称作物理符号系统。用物理符号系统主要是强调所研究的对象是一个具体的物质系统,如计算机的构造系统、人的神经系统、大脑神经元等。所谓符号就是模式;任何一个模式,只要它能和其他模式相区别,它就是一个符号。不同的英文字母就是不同的符号。对符号进行操作就是对符号进行比较,即找出哪几个是相同的符号,哪几个是不同的符号。物理符号系统的基本任务和功能就是辨认相同的符号和区分不同的符号。符号既可以是物理的符号,也可以是头脑中的抽象的符号,可以是计算机中的电子运动模式,也可以是头脑中的神经元的某种运动方式。纸上的文字是物理符号系统,但这是一个不完善的物理符号系统,因为它的功能只是存储符号,即把字保留在纸上。一个完善的符号系统还应该有更多的功能。

图 4.2 给出了物理符号系统的一种框架[529],它由记忆、一组操作、控制、输入和输出构成。它通过感受器输入,输出是确定部位的修改或建立。那么,它的外部行为就由输出组成,它们的产生是输入的函数。大的环境系统加上物理符号系统就形成封闭系统,因为输出变成后面的输入,或者影响后面的输入。物理符号系统的内部状态由它的记忆和控制的状态构成。它的内部行为是由这些内部状态的全部变化构成。

图 4.2 中,记忆是由一组符号结构{$E_1,E_2,\cdots,E_m$}组成的,在整个时间里它们在数量和内容上是变化的。符号结构的内部改变称作表达。为了定义符号结构给出一组抽象符号{$S_1,S_2,\cdots,S_n$}。每种符号结构都具有给定的类型和一些不同的作用{$R_1,R_2,\cdots$},每种作

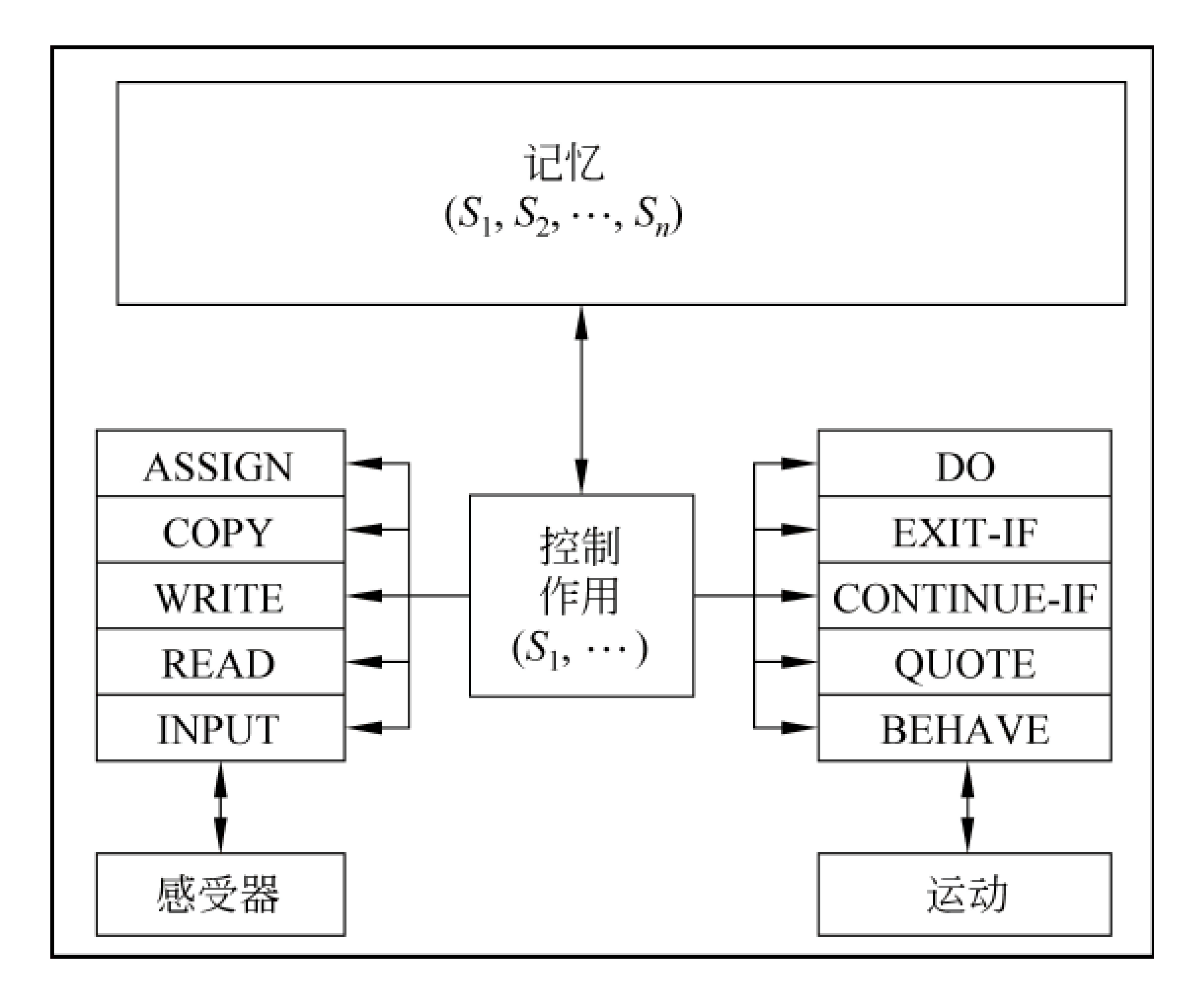



图 4.2　物理符号系统

用包括一个符号。采用显式表示可以写成

$$(\text{Type}: \text{T}\ R_1 : S_1, R_2 : S_2, \cdots, R_n : S_n)$$

若用隐式表示，则写成

$$(S_1, S_2, \cdots, S_n)$$

纽厄尔规定了 10 种操作符[534]。每一个在图中表示一块。这 10 种操作符的功能如下。

(1) 赋值符号(Assign a Symbol)：建立符号与项之间的基本关系。对项赋值，称为存取。符号可以赋给项，而不能赋给表达式。存取一个操作符意味着存取它的输入、输出和唤醒机制。存取给定类型的作用意味着存取作用的符号，这种作用与给定类型的表达式有关，并在那种作用写入新的符号。

(2) 复制表达式(Copy Expression)：将表达式和符号加到系统里，新的表达式是输入表达式准确的复制，即在各种作用中具有完全相同的类型和符号。

(3) 写表达式(Write an Expression)：建立任何规定内容的表达式。它并不建立任何新的表达式，而是修改它的输入表达式。

(4) 写(Write)：在给定的作用下建立一个符号。

(5) 读(Read)：在规定作用下读符号。

(6) 执行序列(Do Sequence)：使系统按规定的序列执行任何动作。

(7) 条件退出和条件继续(Exit-if and Continue-if)：系统行为有条件地继续执行一个序列，或从中退出。

(8) 引用符号(Quote a Symbol)：控制自动地解释被运行的表达式。

(9) 外部行为(Behave Externally)：符号系统可控的外部行为的集合。

(10) 环境输入(Input from Environment)：利用记忆中新建立的表达式将外部环境的输入录入到系统中。

系统行为受控制部分管理。这也是一个机器，它的输入包括操作符。控制行为是由连续的解释组成，这些解释就是运行的表达式。图 4.3 给出了控制操作。控制连续地解释运行的表达式。每个解释结果最终是一个符号。其他影响可以在解释作用时发生，它们也是

解释的一部分。

控制解释运行的表达式，首先判断它是否是一个程序符号结构。这样，控制可以感知结构类型。如果它不是一个程序，那么，解释的结果正好是符号本身。

```
Interpret the active expression:
  If it is not a program:
    Then the result is the expression itself.
  If it is a program:
    Interpret the symbol of each role for that role;
    Then execute the operator on its inputs;
    Then the result of the operation is the result.
Interpret the result:
  If it is a new expression:
    Then interpret it for the some role.
  If it is not a new expression:
    Then use as symbol for role.
```

图 4.3　符号系统控制操作

如果运行的表达式是程序，那么，控制处理执行程序规定的操作。但是，程序中实际符号对操作符的作用和它的输入本身必须能被解释。这些符号不应是操作符和输入，而解释的程序是这些符号。这样，控制解释程序中的每一个符号，直到它最后得到可用作操作符或输入的实际符号。那么，实际它通过发送输入符号到合适的操作符执行操作，唤醒它，并取回操作符产生的结果。

然后，控制解释结果。如果它是一个新表达式，那么就处理解释；如果不是新的表达式，最终得到符号。

我们可以将物理符号系统的功能简化成如下。

（1）输入符号。

（2）输出符号。

（3）存储符号。

（4）复制符号。

（5）建立符号结构：通过找到各种符号之间的关系，在符号系统中形成符号结构。

（6）条件转移：如果在记忆中已经有了一定的符号系统，再加上外界的输入，就可以继续完成行为。

具备上面 6 种处理功能的物理符号系统就是一个完整的物理符号系统。人能够输入符号，如用眼睛看、用耳朵听、用手摸等。通过说话、写字、画图等动作输出。人类可以把输入保存在头脑里，叫做记忆。人通过学习接收信息，然后对符号进行不同的组合，得到新的关系，组成新的符号系统，这是第 4 项和第 5 项功能，即复制和建立新的结构。一个物理符号系统可以根据原来存储的信息，加上当前的输入而进行一系列活动，这就是条件转移。事实上，现代的计算机都具备物理符号系统的这 6 种功能。

1976 年纽厄尔和西蒙提出了物理符号系统假设[534]，说明物理符号系统的本质。主要假设内容如下。

（1）物理符号系统假设：物理系统表现智能行为必要和充分的条件是它是一个物理符号系统。

（2）必要性意味着表现智能的任何物理系统将是一个物理符号系统的示例。

(3) 充分性意味着任何物理符号系统都可以进一步组织表现智能行为。

(4) 智能行为就是人类所具有的那种智能：在某些物理限制下，实际上所发生的适合系统目的和适应环境要求的行为。

由此可见，既然人具有智能，它就是个物理符号系统。人类能够观察、认识外界事物、接受智力测验、通过考试等，这些都是人的智能的表现。人所以能够表现出智能，就是基于他的信息加工过程。这是由物理符号系统的假设得出的第一个推论。第二个推论是，既然计算机是一个物理符号系统，它就一定能表现出智能，这是人工智能的基本条件。第三个推论是，既然人是一个物理符号系统，计算机也是一个物理符号系统，那么我们就能用计算机来模拟人的活动。我们可以用计算机在形式上来描述人的活动过程，或者建立一个理论来说明人的活动过程。

1981 年纽厄尔以物理符号系统为中心，以纯认知功能为基础建立了纯认知系统模型[530]，如图 4.4 所示。

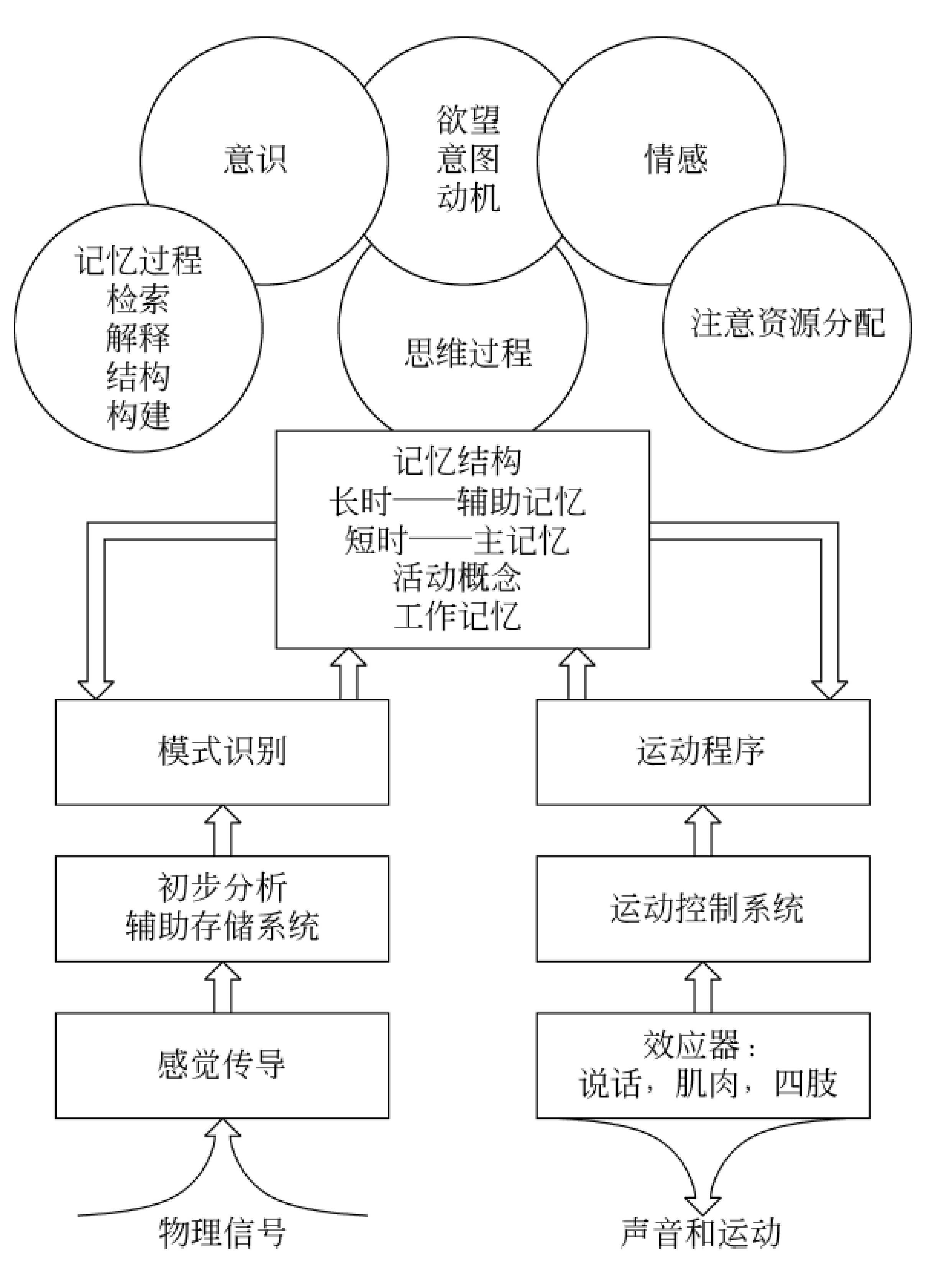



图 4.4 纯认知系统模型

## 4.4 ACT 模型

美国心理学家安德森于 1976 年提出系统的整合理论与人脑如何进行信息加工活动的理论模型，简称 ACT(Adaptive Control of Thought)模型[25]，原意为“思维的自适应控制”。

安德森将人类联想记忆模型(HAM)与产生式系统的结构相结合,模拟人类高级认知过程的产生式系统,在人工智能的研究中有重要意义。ACT 模型着重强调高级思维的控制过程,已经发展有下列版本的 ACT 系统。

1978:ACT*。

1993:ACT-R。

1998:ACT-R 4.0。

2001:ACT-R 5.0。

2004:ACT-R 6.0。

1983 年,安德森在《认知结构》一书中从心理加工活动的各个方面对其基本理论进行阐述,他所提出的 ACT 产生式系统的一般框架由 3 个记忆部分组成:工作记忆、陈述性记忆和产生式记忆(如图 4.5 所示)[26]。

- 陈述性记忆,也就是一个具有不同激活强度且由相互连接的概念所构成的语义网络。
- 程序性记忆,即一系列产生式规则的程序性记忆。
- 工作记忆,它包含当前被激活的那些信息。

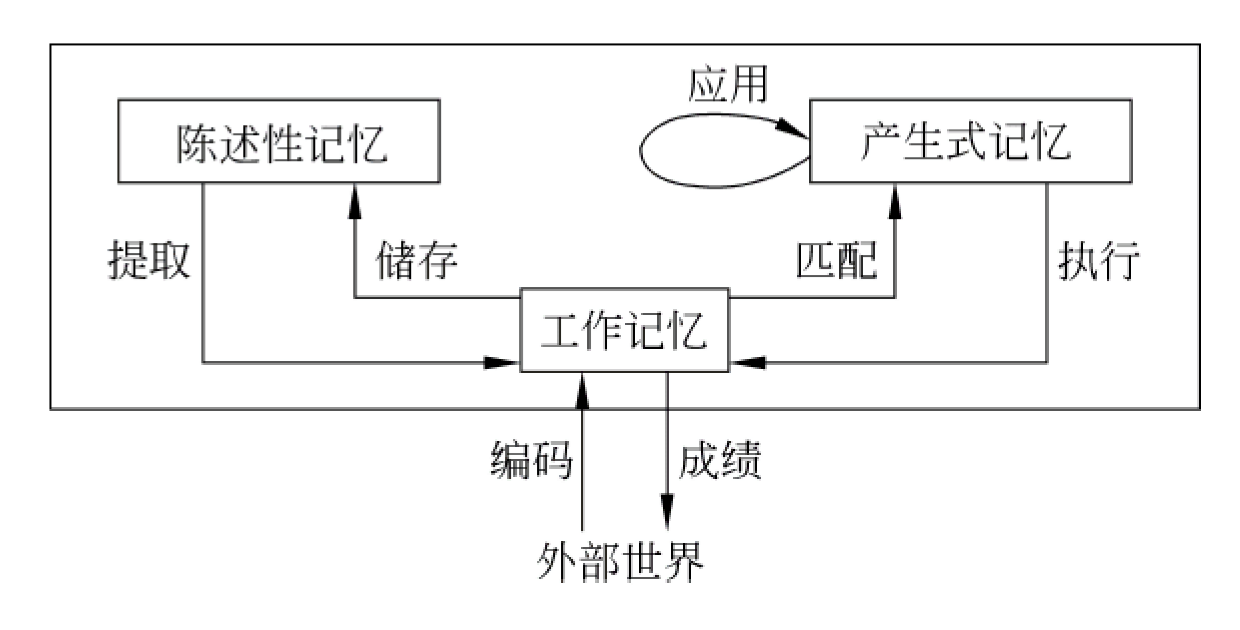



图 4.5 ACT 的系统结构

陈述性知识是以组块为单位来表征的。组块类似于图式结构,每一组块能对小组知识进行编码。陈述性知识是能够被报告的,并且不与情境紧密关联,而程序性知识通常是不能被表达的,是自动运用的,而且是有针对性地被应用到特定情境中。被试可通过多种方法把信息存储在陈述性记忆中并提取出来。匹配过程是把工作记忆中的材料与产生式的条件相对应,执行过程是把产生式匹配成功所引起的行动送到工作记忆中。在执行前的全部产生式匹配活动也称为产生式应用,最后的操作由工作记忆完成,这些规则就能够得到执行。通过“应用程序”,程序性记忆能被运用到其自身加工之中;通过检查已经存在的产生式,被试能学习到新的产生式。在最大限度上,安德森把技能获得解释成为知识编译,也就是实现陈述性知识到程序性知识的转变过程。知识编译具有两个子过程:程序化与合成。

程序化是指把陈述性知识转化成程序性知识或产生式知识的过程。问题解决者开始时常根据书本知识来解决诸如数学或编程这样的问题。在尝试解决问题的过程中,新手就会用爬山法和手段——目的分析这样的弱方法的组合去产生许多子目标,并且产生陈述性知识。当多次解决问题中的某一事件时,在一个特别的情景下,一段特别的陈述性知识就会被反复地提取出来。这个时候,一个新的产生式规则就形成了。在应用中可以学习到新的产生式,这表明依据 ACT 理论,过程学习是“做中学”的。过程性知识可被描述成一个模式(产生式的 IF 部分),而被执行的动作被描述为动作(产生式的 THEN 部分)。这种陈述性

知识到程序性知识的转化过程,会同时导致被试言语化加工的减少。与此相关的是,问题解决行为的自动化程度会有所提高。

在 ACT-R 中,学习是根据微小知识单元的增长和调整而实现的。这些知识能够组合起来产生复杂的认知过程。在学习过程中,环境扮演了重要的角色,因为它建立了问题对象的结构。这个结构能协助进行组块学习,并促进产生式规则的形成。这一步的重要性在于,它重新强调了作为理解人类认知的关键步骤的分析环境的本质特征的重要性。但是,自从行为主义消亡、认知革命兴起以来,这一点却被忽视了[27]。

## 4.5　SOAR 模型

20 世纪 50 年代末,在对神经元的模拟中提出了用一种符号来标记另一些符号的存储结构模型,这是早期的组块(Chunks)概念。在象棋大师的头脑中就保存着在各种情况下对奕经验的存储块。20 世纪 80 年代初,纽厄尔和罗森勃卢姆(P Rosenbloom)认为,通过获取任务环境中关于模型问题的知识,可以改进系统的性能,组块可以作为对人类行为进行模拟的模型基础。通过观察问题求解过程,获取经验组块,用其代替各个子目标中的复杂过程,可以明显提高系统求解的速度。由此奠定了经验学习的基础。1987 年,纽厄尔和莱德(J Laird)、罗森勃卢姆提出了一个通用解题结构 SOAR[394],希望能把各种弱方法都实现在这个解题结构中(图 4.6)。

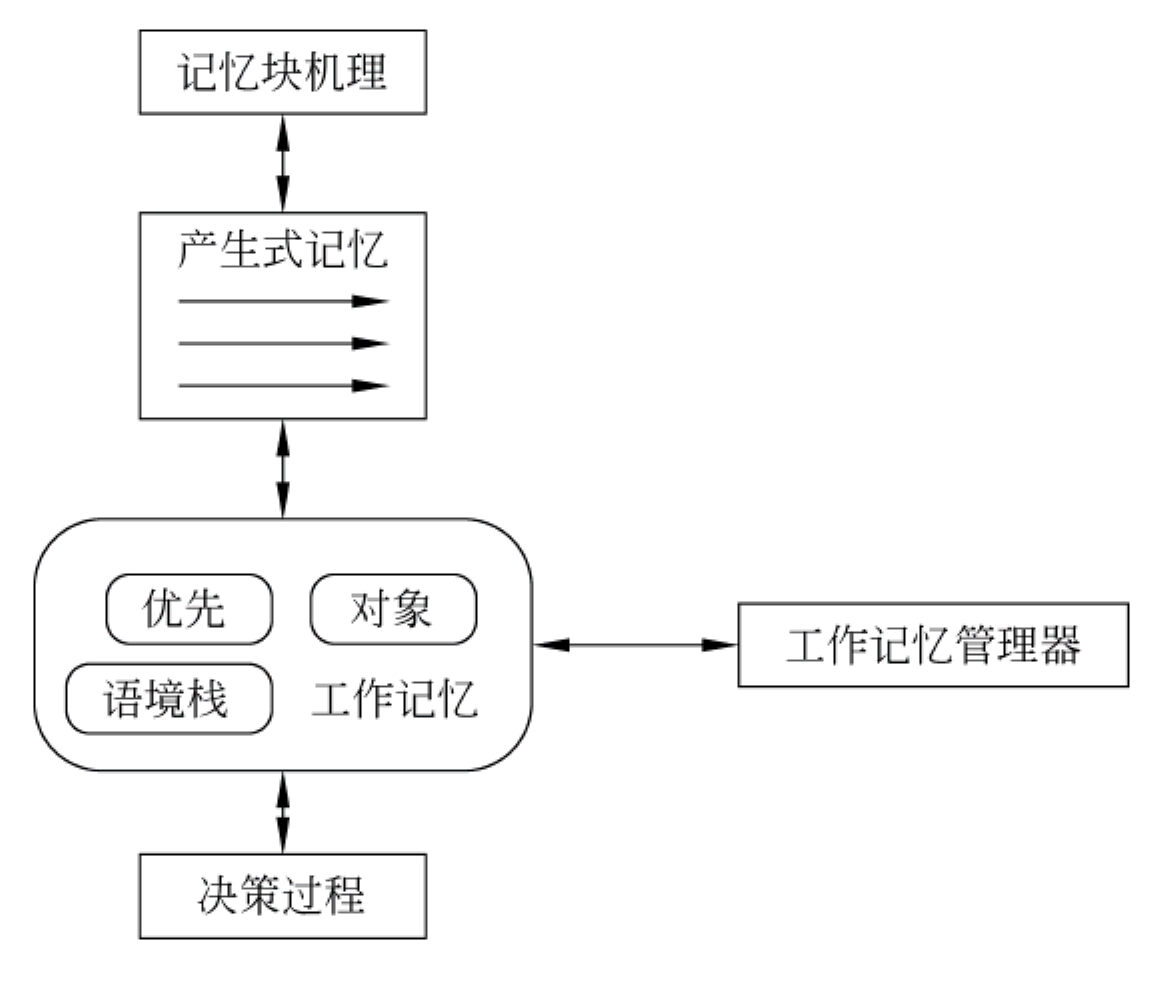



图 4.6　SOAR 的框图

SOAR 是 State, Operator and Result 的缩写,即状态,算子和结果之意,意味着实现弱方法的基本原理是不断地用算子作用于状态,以得到新的结果。SOAR 是一种理论认知模型,它既从心理学角度,对人类认知建模,又从知识工程角度,提出一个通用解题结构。SOAR 的学习机制是由外部专家的指导来学习一般的搜索控制知识。外部指导可以是直接劝告,也可以是给出一个直观的简单问题。系统把外部指导给定的高水平信息转化为内部表示,并学习搜索组块。

产生式记忆器和决策过程形成处理结构。产生式记忆器中存放产生式规则,它进行搜

索控制决策分为两个阶段。第一阶段是详细推敲阶段,所有规则被并行地用于工作记忆器,判断优先权,决定哪部分语境进行改变,怎样改变。第二阶段是决策阶段,决定语境栈中要改变的部分和对象。

SOAR 中的所有成分统称为对象,这些成分包括状态、状态空间、算子和目标。所有这些对象都存放在一个叫 Stock 的库中,因此,库中也划分为这么四个部分。另有一个当前环境,也同样地分为四个部分。其中每个部分最多存放库中相应部分的一个元素。例如,当前环境的状态部分可以存放库中的一个状态,称为当前状态,等等。当前环境的一个部分也可以不存放任何东西,此时认为该部分无定义。例如,若没有任何算子可作用于当前状态,则当前环境的算子部分成为无定义。为什么要把状态和状态空间分成两个独立的部分呢?这是因为在解题过程中有时可能需要改变问题的形式,从而从一个状态空间转移到另一个状态空间。

在 SOAR 问题求解过程中,如何利用知识空间的知识非常重要。利用知识控制 SOAR 运行的过程,大体上是一个分析—决策—行动的三部曲。

### 1. 分析阶段

输入:库中的对象。

任务:从库中选出对象加入当前环境。

增加有关当前环境中对象的信息角色。

控制:反复执行,直至完成。

### 2. 决策阶段

输入:库中的对象。

任务:赞成,或反对,或否决库中的对象。选择一个新的对象,用它取代当前环境中的同类对象。

控制:赞成和反对同时进行。

### 3. 执行阶段

输入:当前状态和当前算子。

任务:把当前算子应用于当前状态。

如果因此而产生一个新状态,则把新状态加入库中,并用它取代原来的当前状态。

控制:这是一个基本动作,不可再分。

SOAR 系统运行过程中,在分析阶段,任务是尽量扩大有关当前对象的知识,以便在决策阶段使用。决策阶段主要是进行投票,投票由规则来做,它可以看成是同时进行的,各投票者之间不传递信息,不互相影响。票分赞成、反对和否决三种。每得一张赞成票加一分,得一张反对票减一分,凡得否决票即绝对无中选的可能。在执行阶段,如果当前环境的每个部分都有定义,则用当前算子作用于当前状态。若作用成功,则用新状态代替旧状态,算子部分成为无定义,重新执行分析阶段。

分析和决策阶段是通过产生式系统来实现的,产生式的形式是

$$C_1 \land C_2 \land \cdots \land C_n \rightarrow A$$

条件 $C_i$ 是否成立取决于当前环境和库中的对象情况，$A$ 是一个动作，它的内容包括增加某些对象的信息量和投票情况等。

每当问题求解器不能顺利求解时，系统就进入劝告问题空间请求专家指导。专家以两种方式给以指导。一种是直接指令方式，这时系统展开所有的算子以及当时的状态。由专家根据情况指定一个算子。指定的算子要经过评估，即由系统建立一个子目标，用专家指定的算子求解。如果有解，则评估确认该算子是可行的，系统便接受该指令，并返回去求证用此算子求解的过程为何是正确的。总结求证过程，从而学到使用专家劝告的一般条件，即组块。

另一种是间接的简单直观形式，这时系统先把原问题按语法分解成树结构的内部表示，并附上初始状态，然后请求专家劝告。专家通过外部指令给出一个直观的简单问题，它应该与原问题近似，系统建立一个子目标来求解这个简单问题。求解完后就得到算子序列，学习机制通过每个子目标求解过程学到组块。用组块直接求解原问题，不再需要请求指导。

SOAR 系统中的组块学习机制是学习的关键。它使用工作记忆单元来收集条件并构造组块。当系统为评估专家的劝告，或为求解简单问题而建立一个子目标时，首先将当时的状态存入工作记忆单元 w-m-e。当子目标得到解以后，系统从 w-m-e 中取出子目标的初始状态，删去与算子或求解简单问题所得出的解算子作为结论动作。由此生成产生式规则，这就是组块。如果子目标与原问题的子目标充分类似，组块就会被直接应用到原问题上，学习策略就把在一个问题上学到的经验用到另一个问题上。

组块形成的过程可以说是依据对于子目标的解释而请示外部指导，然后将专家指令或直观的简单问题转化为机器可执行的形式，这运用了传授学习的方法。最后，求解直观的简单问题得到的经验（即组块）被用到原问题，这涉及类比学习的某些思想。因此可以说，SOAR 系统中的学习是几种学习方法的综合应用。

## 4.6　心智社会

美国《连线》（*Wired*）杂志创始主编凯利（K Kelly）于 1994 年写了《失控：全人类的最终命运和结局》一书[362]，全书有 24 章，其中第三章是“有心智的机器”。凯利指出“自然进化强调我们是猿类；而人工进化则强调我们是有心智的机器……作为人类，当我们知道自己是这颗蓝色星球上枝繁叶茂的生命之树上的一根枝条时，我们就找到了精神的家园。也许将来某一天，当我们知道自己是层积在绿色生命之上的复杂机器中的一根纽带时，我们将进入精神的天堂。自旧的生命系统中诞生出新生命的庞大网络，人类则成为其中一个华丽的节点。”凯利赞美明斯基“他不遗余力地鼓吹把人脑的智能下载进计算机”。明斯基主张“我们要给机器赋予智慧，让他们有自我意识”[362]。

明斯基于 1985 年出版了《心智社会》一书[504]。他在这本书中指出，智能并非存在于中央处理器中，而是由许多具有专门用途、彼此紧密联结的机器的集体行为中产生的。明斯基指出：心智是由许多称作智能体（Agent）的小处理器组成；每个智能体本身只能做简单的任务，它们并没有心智；当智能体构成社会时，就产生智能。从脑部高度关联的互动机制中，涌现出各种心智现象。丹尼特（D Dennett）也认为“有许多微不足道的小东西，本身并没有什么意义，但意义正是通过其分布式交互而涌现出来的”。

美国著名机器人专家布鲁克斯(R A Brooks)的移动机器人实验室,开发出来的一套分布式控制方法:

(1) 先做简单的事。

(2) 学会准确无误地做简单的事。

(3) 在简单任务的成果之上添加新的活动层级。

(4) 不要改变简单事物。

(5) 让新层级像简单层级那样准确无误地工作。

(6) 重复以上步骤,无限类推。

这套方法是"众愚成智"的体现。

明斯基在《心智社会》中提出,把意识移植到机器内将可能实现。至少从20世纪30年代起人们就知道,人脑中存在着电子运动,也就是说,人的记忆甚至个性都可能是以电子脉冲的形式存在的。于是从原则上可能通过某种电子机械设备测定这些脉冲并把它们在另外一个媒介,如记忆库中复制出来。由此一来,记忆中的"我"作为本质上的"我",可以在计算机里保存,记忆可以被复制、移植和数字化运作,成为真实自我的数字展现。这样,即使在计算机里,"我"仍可以得到同以前完全相同的体验。对作为自我意识的数字化除了可以设想复制后移出外,还可以有一种反向的过程,就是将体外的自我意识——可以是他人的自我意识,也可以是经过机器加工处理后的自我意识——移入自我的头脑,从而形成新的自我意识。

## 4.7 LIDA

富兰克林(Stan Franklin)和他的同事们提出了LIDA[245],它是一个基于意识的全局工作空间理论[42]构建的概念性心智模型。LIDA计算架构来自LIDA心智模型,以LIDA认知周期为基础。每一个智能体[233]、人类、动物还是人工智能系统,必须都必须频繁地采集(感知)环境信息,处理(理解)输入,并选择合适的响应(动作)。智能体的"生命"被看作是由认知周期的连续序列组成。每个周期构成一个知觉、处理和行动的单元。

### 1. 全局工作空间理论

全局工作空间方法的目的在于详细说明大脑意识活动在认知活动中的角色和作用[41]。图4.7给出了全局工作空间意识模型[42]。用戏院来比喻全局工作空间方法是最为贴切的。观众作为戏院中无意识的处理器,接受舞台上意识焦点的广播。对于这个焦点的控制相当于选择性的注意。在后台,无意识的连接系统形成和导向有意识的内容。全局工作区间方法是可检验假设的严格集合,这个比喻只是为了说明它的一些基本特征[43]。

全局工作空间方法是基于有意识的处理过程的强有力证据之上的,有意识的处理过程结合了"黑板结构"这个人工智能的概念,这个结构结合不同认知来源以在复杂嘈杂含糊的环境中鉴别出一个声音信号。这样多噪声含糊的信号在人们平时观察、思考、动机计划和控制等情况中是常见的。基于大量有意识和无意识活动进行比较的实验,人们提出了经验性的总结[41]。这表明有意识的活动是和有限容量的处理过程强烈相关的,表现出有意识内容

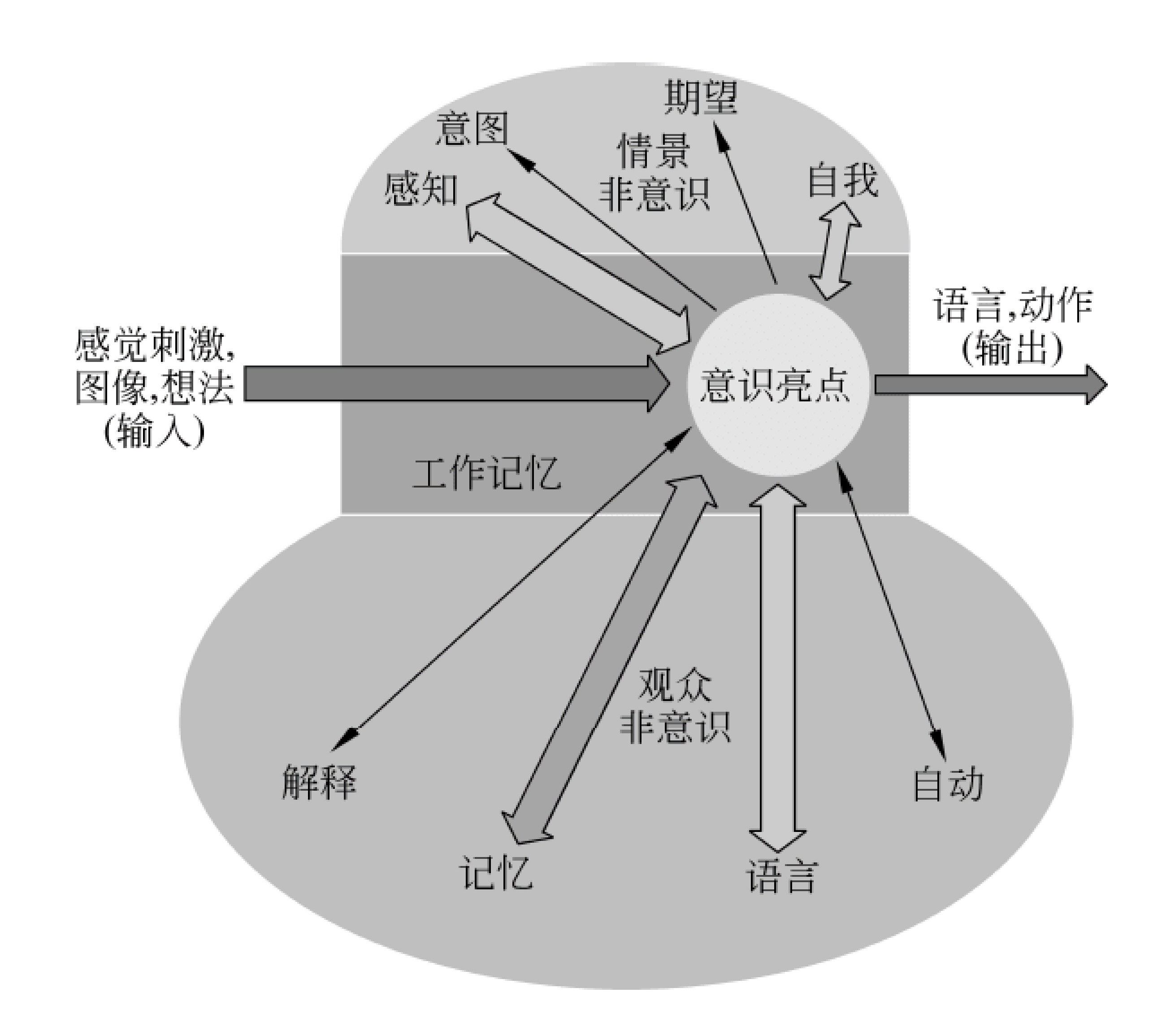



图 4.7　全局工作空间意识模型

的内部相容性和低计算效率。

第一个全局工作空间结构是纽厄尔和他的合作伙伴一起提出的,用来分辨在嘈杂环境中的词语声音。这是一个困难的挑战,因为空间会把反射和背景噪声加到一个已分辨出来的声音信号上,这会使得标准的模式识别算法很大程度上效率低。纽厄尔等人的全局工作区间模型 Hearsay 解决了这个问题,正如人们必须生存和繁殖一样。因此全局工作空间系统表现出一个很大的功能性优点,弥补了前面问题的缺点。

全局工作空间方法假设人类的认知过程是有相当多几乎全是无意识的相当小的专门处理过程来执行的。尽管今天看起来已经是老生常谈,但是大脑中广泛分布的专门处理的概念在刚提出时还是有很大争议的。不同的处理需要竞争使用全局工作空间的使用通道。这种有限容量的全局工作空间方法用于将信息由竞争胜出的处理部件传递到所有无意识的处理部件上,目的是为了采纳更多资源加入已处理异常的和高优先级的输入,解决目前存在的问题。从这个角度来看意识允许我们处理那些常规的无意识处理不能有效处理的异常或者有挑战性的情况。有意识的认知能够通过授予通向那些不可预知但是又必需的知识源的通道来解决人工智能和机器人学科中遇到的"关联问题"。在默认的情况下,意识会提供一个查找函数来发现潜在的危险和机遇,所以在意识的内容和重要的感知输入之间存在特殊的紧密联系。外部的感觉能通过感知到的视觉表象、内心语言和内心感觉行为等途径来自发模拟出来。这些内源性"感觉"已经被证明可以调动因为类似外部活动而变得活跃的皮层及皮下区域。

有意识的内容总是被无意识的内容引导和约束:目标内容、感知的内容、概念上的内容和共享环境内容。每个内容本身就是一个处理的结合,尽管这些内容本身是无意识的,但是它们形成有意识的处理过程。例如,无意识的空间认知会被要求说明意识到的观察目标的方向。在全局工作空间方法中,学习活动普遍是由有意识的部分引起的。当无意识的处理器被有意识的内容要求执行时,难以报告的规律生成问题解决时,隐性的学习发生了。像语言学习任务很大程度上都是隐性的,但它们主要是由有意识的输入引起的。

### 2. 智能数据分析认知结构

智能数据分析是为美国海军执行专家预定要求的任务而开发的智能软件[233]。当一个海员服役期满,他就被分配到一个新的地方。分配须综合考虑海员的选择、海军的要求以及大量的规则条例,这个任务的处理过程被称作分配。海军需要聘用 300 个全职推销员来完成这些分配。这款软件使得这个工作完全变成自动化处理,已经投入使用,效果和人工工作相当。

这款软件包括一个计算的部件,已经在执行和运转,一款更高性能的智能数据分析软件也将要投入运转。除了可以合并全局工作空间方法实现和有意识的认知等价的功能以外,智能数据分析还可以作为很多心理学方法的工具。在复杂和动态变化的环境中智能体会频繁反复进行抽样并做出响应。大量的人类认知活动都是通过意识的内容、不同的记忆系统以及选择之间的快速交互作用来执行的,把这些活动称为认知周期[46]。与认知周期相一致的发现已经被很多神经系统科学家报道。认知周期可以同时发生,产生并行的响应,但是有意识的内容只能串联出现。智能数据分析模型表明有意识的认知只能以不连贯的序列出现,被非常短的无意识内容所分离。

智能数据分析是实现全局工作空间方法的高精度仿真的重要的一步,包括了如何指定全局功能工作空间方法仅仅作为黑盒子处理的特殊知识源。人们正在开发一个领域独立版本的智能数据分析以处理更大范围的任务。智能数据分析的目标是成为一个能力和人类有意识的内容相一致的智能体。

智能数据分析的认知周期可以分为如下 9 步(如图 4.8 所示)[46,732]。

(1) 输入的感知刺激因素被前意识知觉所过滤,加进意图,产生感知对象。

(2) 目前的感知对象被移至前意识的工作记忆,和此前的循环还没有衰减的感知对象一起加入更高级别感知的结构当中。

(3) 当前工作记忆的结构会诱发瞬时情景记忆和陈述性记忆,产生储存在长期工作记忆的局部联系。

(4) 长期工作记忆的内容的组合为意识展开竞争,使得意识可能把系统资源用在最合适最紧急最重要的任务成分上。

(5) 有意识的广播按全局工作空间方法产生,使得不同形式的学习以及内部资源的补充成为可能。基于一些经验性资源,假设广播需要 100ms 的时间。

(6) 程序性记忆需要响应有意识广播的内容。

(7) 产生其他(无意识)的响应方案,例如行为选择机制的自我复制,捆绑变量,激活传递等。

(8) 行为选择机制为该认知循环选择一个响应。

(9) 智能数据分析对内部或者外部环境产生响应。

LIDA 认知周期可分为三个阶段,理解阶段、注意(意识)阶段和动作选择阶段(如图 4.8 所示)。它开始于左上角,大致以顺时针流动。理解阶段开始于刺激输入,激活感觉记忆中的初级特征检测器。输出信号被发送到感觉联想记忆中,在那里更高层次的功能探测器用于更抽象的实体,如对象、类别、行动、事件等的检测。所产生的知觉移动到工作区,在那里产生本地联系的短暂情景记忆和陈述性记忆会被做线索标记。这些本地联系与知觉

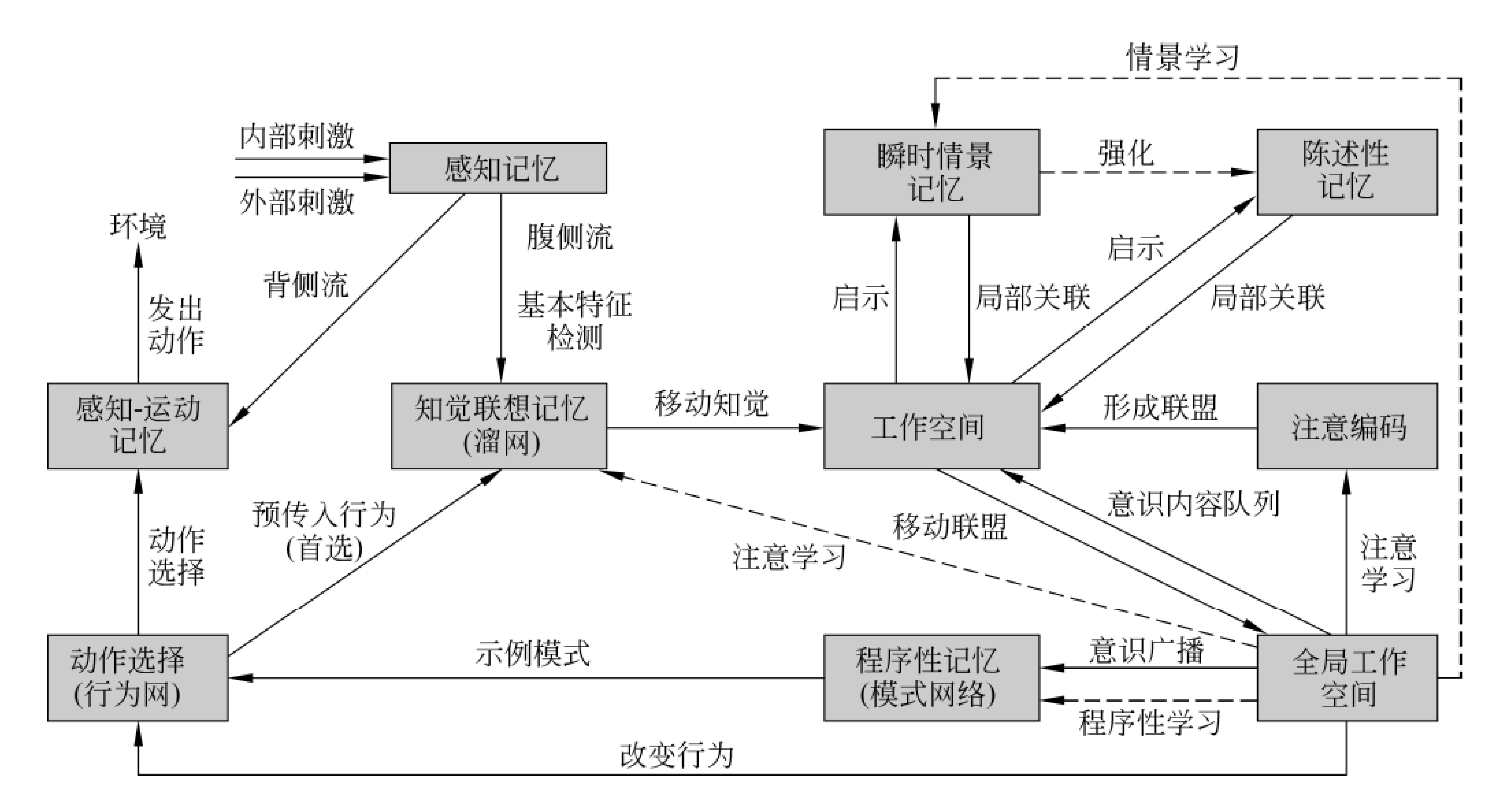



图 4.8 LIDA 认知周期

结合,产生当前情景模型,用以表示智能体对当前正在发生的事情的理解。

注意编码器通过当前情景模型中选定部分之间形成联接,开始了注意阶段,并把形成的联接转移到全局工作空间。全局工作空间的竞争随后选出最突出、最相关、最重要和最紧迫的联接,它们的内容就成为意识的内容。然后,这些意识的内容被广播到全空间,启动行动的选择阶段。LIDA 认知周期的行动选择阶段也是一个学习阶段,这个学习阶段中会有几个处理操作并行进行。新的实体和联合,以及加强过的旧有意识会随着意识发布到知觉联想记忆而发生。意识广播中的事件将瞬态情景记忆的编码为新的记忆。可能采取的行动计划,连同它们的背景和预期的结果,从意识发布提取到程序记忆中。旧的计划得到了加强。与所有学习过程并行发生的、可能的行动主体使用意识内容从程序记忆中再被使用。每个这样的副本是与其绑定的变量一起实例化,并传送到行动的选择环节,竞争的结果就是此轮认知周期选择的行为。选定的行为触发感觉运动记忆,产生一个合适的算法执行此行为。认知周期结束。

工作区内部结构是由各种不同的输入缓冲区和三个主要模块组成:当前情景模型、暂存器和意识内容队列[732]。当前情景模型是一个存储代表实际的当前内部和外部事件的结构。结构构建编码器负责创建使用工作区的各种子模型中的元素的结构。暂存器是在工作区的辅助空间,在这里结构构建编码器可以构建可能的结构,然后将它们转移到当前的情景模式。队列意识的内容存储着持续的几个广播的内容,使 LIDA 模型理解和操作与时间有关的概念[732]。

### 3. LIDA 软件框架

LIDA 软件框架的主要目标是提供一个 LIDA 模型的通用实现,容易定制化具体问题或领域。这样一个通用的、可定制的实施是用 Java——一个强大和成熟的面向对象的语言实现的。

LIDA 框架定义了几种数据结构和程序(算法),并由一些小块组成。在 LIDA 模型中,其主要成分为模块,是用相互关联的元素代表。另一个主要成分是控制框架中的所有进程的执行的任务管理器。由微小的、功能强的处理器处理的这些过程就叫做 LIDA-任务。LIDA-任务可以以一种对用户几乎透明的方式通过 LIDA-任务管理器的单线程实施。节点结构是构成一个框架中的主要数据结构的核心要素。最后,提供支撑性的软件工具,如定制化的图形用户界面、日志记录和系统加载器。系统加载器解析带有应用程序定义和参数的 XML 文件。

对 LIDA 认知模型的每一个主要部件,在框架里都定义了一个组件。LIDA 框架中的大部分组件是领域独立。每个组件,框架都提供一个默认的实现。例如,情节记忆是使用稀疏分布的内存和由一个大量修正的行为网络的行动选择组件一起实现的。开发人员可以直接使用这些实现并定制一些参数。然而还有一些模块是属于特定领域的,必须由用户去指定,特别是感觉记忆和感觉运动记忆。不过,用户可以使用框架提供的这些默认的组件去扩展实现自己领域特定的实现。

组件之间通信是通过“监听器”组件实现。它接收到来自另一个“生产者”组件的信息,向生产者注册一个监听器的角色。每当生产者有信息要发送,它都会把信息传送给所有在它这里注册过的监听器组件。框架中会使用大量这种监听器组件的实例。每个监听器类型实现时带有自己的接口。一个组件可以注册为其他组件的监听器。同时监听器可同时是其他组件的生产者或者监听者。这种模式有几个好处,主要是,监听器和生产者不需要知道对方的内部构造和具体实现,他们只需要满足特定的监听器接口即可。图 4.8 中的箭头在框架中是作为监听者来实现的。

框架的另一个重要部分是节点的数据结构。一个节点结构是一个图结构,包含节点和它们之间的连接。它构成了许多框架组件中的主要数据表示。模块使用节点结构表示它们的内部数据,而在框架中也会使用其他的表示形式。节点结构极大地有助于创建图结构,因为它们管理着对低级的特定节点的操作需要,诸如添加、删除或检索和连结。这些图用来作为对象、行动和事件的概念性表示,是 LIDA 模型中的基本数据表示。

模块为完成特定功能要执行几个任务。框架提供 LIDA 任务和小流程的封装。每个 LIDA 任务有一个算法、执行时间和状态。一个模块可以创建几个 LIDA 的任务来帮它执行功能。每个 LIDA 任务可以运行一次或重复执行。LIDA 任务的执行被委派给 LIDA 任务管理器——这是框架的重要部分,负责应用程序所有任务的调度和执行。它维护一个线程池,因此可以同时执行多个任务。任务管理器维护一个任务队列用以调度要执行的 LIDA 任务。

LIDA 软件框架允许创建基于 LIDA 模型的新智能软件。它的设计和实施,旨在简化这一过程,并允许用户关注于实现应用程序的细节,隐藏模型的通用部件的复杂性。它也保留了能简化创建复杂结构的良好程序算法。它允许以多种方式和与模块之间的耦合程度低层次的定制来实现高层次的抽象。还提供了一个可定制的图形用户界面和日志记录的补充工具。结果就是,这是一个开发基于 LIDA 应用程序的强大的和可定制的工具,可能同样也适用于开发其他应用程序。

# 4.8　CAM 心智模型

在人的心智中，记忆和意识是最为重要的两个部分。其中记忆存储各种重要信息和知识，意识让人有自我的概念，能根据需求、偏好设定目标，并根据记忆中的信息进行各种认知活动。我们主要基于记忆和意识构建了 CAM(Consciousness and Memory)心智模型[688]。如图 4.9 所示，CAM 包括 10 个主要模块，简单介绍如下。

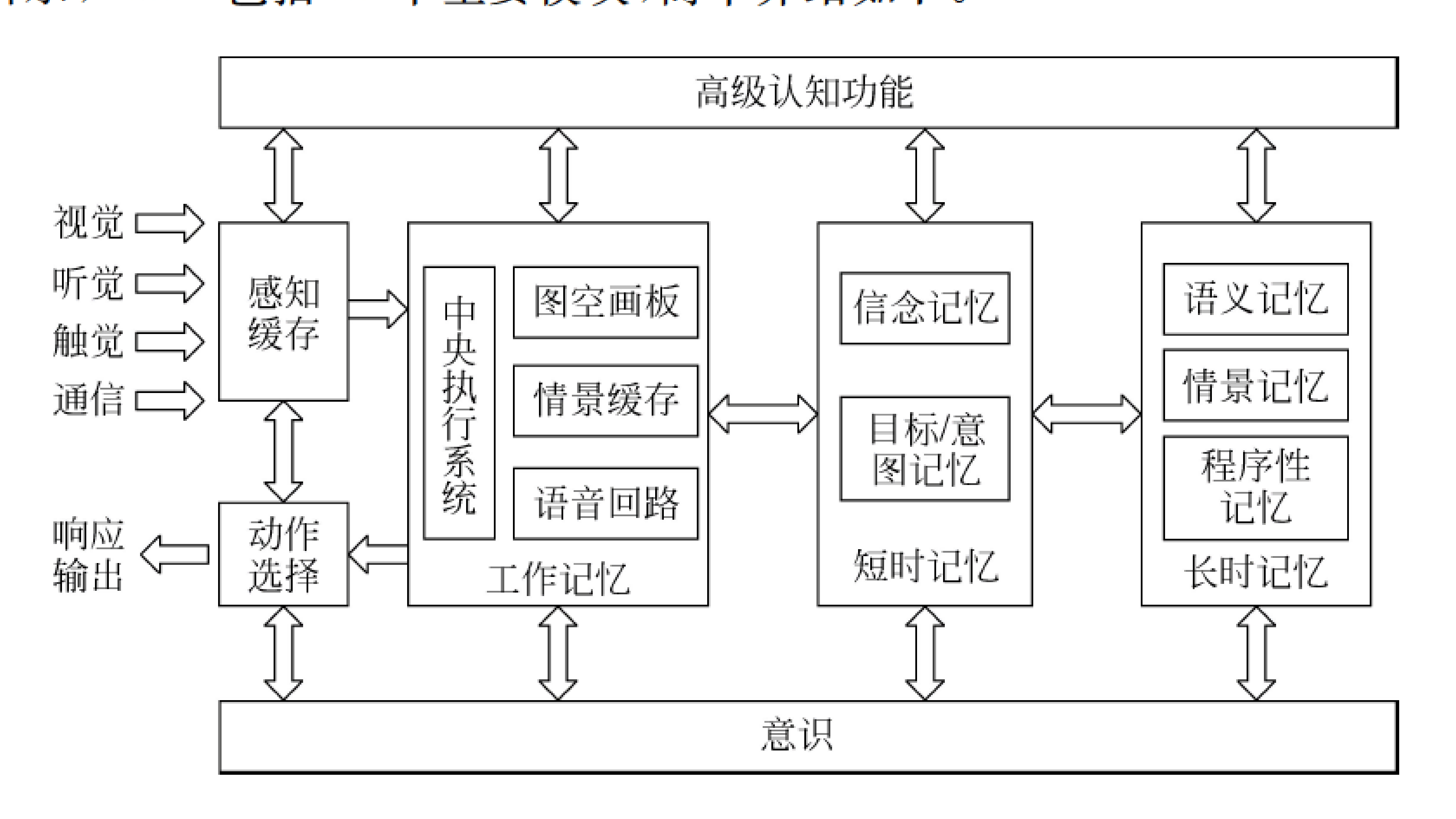



图 4.9　CAM 的系统结构

## 1. 视觉

人的感觉器官包括视觉、听觉、触觉、嗅觉、味觉。在 CAM 模型中重点考虑视觉和听觉。视觉系统使生物体具有视知觉能力。它使用可见光信息构筑机体对周围世界的感知。根据图像发现周围景物中有什么物体和物体在什么地方的过程，也就是从图像中得到对观察者有用的符号描述的过程。视觉系统具有将外部世界的二维投射重构为三维世界的能力。需要注意的是，不同物体所能感知的可见光处于光谱中的不同位置。

外界的物体在视网膜成像时，实际上是光线这个刺激因素被视网膜的感光细胞(视杆细胞和视锥细胞)转变为电信号，后者经视网膜内双极细胞传到神经节细胞形成神经冲动，即视觉信息。视觉信息再经视神经传向脑。双极细胞可看成是视觉传导通路的第 1 级神经元，神经节细胞是第 2 级神经元。很多神经节细胞发出的神经纤维组成较粗大的视神经，视神经在眼球的后端离开眼球向后进入颅腔，这时，左右侧的视神经发生了交叉，交叉的部位我们称之为视交叉。视束在大脑底面向后连于外侧膝状体。外侧膝状体是一个重要的视觉信息传导的中间站，其中含有的第 3 级神经元，它们发出的大量纤维组成所谓的视辐射，视辐射的纤维最后投射到大脑枕叶的视觉中枢，即视皮质。视觉信息只有传到脑的视觉皮层并经过处理、分析，才能最后形成主观的视觉感受。

视觉皮层是指大脑皮层中主要负责处理视觉信息的部分，位于大脑后部的枕叶。人类的视觉皮层包括初级视皮层(V1，亦称纹状皮层)以及纹外皮层(V2，V3，V4，V5 等)。初级视皮层位于 17 区。纹外皮层包括 18 区和 19 区。

初级视皮层(V1)的输出信息送到两个渠道,分别成为背侧流和腹侧流。背侧流起始于V1,通过V2,进入背内侧区和中颞区(MT,亦称V5),然后抵达顶下小叶。背侧流常被称为"空间通路",参与处理物体的空间位置信息以及相关的运动控制,例如眼跳和伸取。腹侧流起始于V1,依次通过V2,V4,进入下颞叶。该通路常被称为"内容通路",参与物体识别,例如面孔识别。该通路也与长时记忆有关。

### 2. 听觉

人们之所以能听到声音、理解言语,是依赖于整个听觉通路的完整性,它包括外耳、中耳、内耳、听神经及听觉中枢。听觉通路在中枢神经系统之外的部分称为听觉外周,在中枢神经系统内的部分称为听觉中枢或中枢听觉系统。听觉中枢纵跨脑干、中脑、丘脑的大脑皮层,是感觉系统中最长的中枢通路之一。

声音信息自周围听觉系统传导至中枢听觉系统,中枢听觉系统对声音有加工、分析的作用,像感觉声音的音色、音调、音强、判断方位。还有专门分化的细胞,对声音的开始和结束分别产生反应。传到大脑皮层的听觉信息还与大脑中管理"读""写""说"的语言中枢相联系,有效完成我们经常用到的读书、写字、说话等功能。

### 3. 感知缓存

感知缓存又称感觉记忆或瞬时记忆,是感觉信息到达感官的第一次直接印象。感知缓存只能将来自各个感官的信息保持几十到几百毫秒。在感知缓存中,信息可能受到注意,经过编码获得意义,继续进入下一阶段的加工活动,如果不被注意或编码,它们就会自动消退。

各种感觉信息在感知缓存中以其特有的形式继续保存一段时间并起作用,这些记忆形式就是视觉表象和声音表象,称视象和声象。表象可以说是最直接、最原始的记忆。表象只能存在很短的时间,如最鲜明的视象也不过持续几十秒钟。感觉记忆具有下列特征。

(1) 记忆非常短暂。

(2) 有能力处理像感受器在解剖学和生理学上所能操纵的同样多的物质刺激能量。

(3) 以相当直接的方式进行信息编码,瞬时保存感觉器官传来的各种信号。

### 4. 工作记忆

工作记忆由中枢执行系统、视觉空间画板、语音回路和情景缓存构成。中枢执行系统是工作记忆的核心,负责各子系统之间以及它们与长时记忆的联系、注意资源的协调和策略的选择与计划等。视觉空间画板主要负责储存和加工视觉空间信息,可能包含视觉和空间两个分系统。语音回路负责以声音为基础的信息的储存与控制,包含语音储存和发音控制两个过程,能通过默读重新激活消退的语音表征防止衰退,而且还可以将书面语言转换为语音代码。情景缓存记忆跨区域的连接信息,以便按时间次序形成视觉、空间和口头信息的集成单元,例如一个故事或者一个电影景物的记忆。情景缓存也联系长时记忆和语义的内容。

### 5. 短时记忆

短时记忆存储信念、目标和意图的内容。它们响应迅速变化的环境条件和智能体的运作方案。知觉的短时记忆存储相关物体的关系编码方案和经验期望编码的预先知识。

### 6. 长时记忆

长时记忆的信息保持时间长，容量大。长时记忆按其内容不同，可分为语义记忆、情景记忆、程序性记忆。

(1) 语义记忆存储的信息是词、概念、规律，以一般知识作参考系，具有概括性，不依赖于时间、地点和条件，不易受外界因素干扰，比较稳定。

(2) 情景记忆的信息是个人亲身经历的、发生在一定时间和地点的事件(情景)的记忆，容易受各种因素的干扰。

(3) 程序性记忆是指关于技术、过程或“如何做”的记忆。程序性记忆通常较不容易改变，但可以在不自觉的情况下自动行使，可以只是单纯的反射动作，或是更复杂的一连串行为的组合。程序性记忆的例子包括学习骑脚踏车、打字、使用乐器或是游泳。一旦内化，程序记忆是可以非常持久的。

### 7. 意识

意识(Consciousness)是一种复杂的生物现象，哲学家、医学家、心理学家对于意识的概念各不相同。从智能科学的角度，意识是一种主观体验，是对外部世界、自己的身体及心理过程体验的整合。意识是一种大脑本身具有的“本能”或“功能”，是一种“状态”，是多个脑结构对于多种生物的“整合”。在心智模型 CAM 中，意识是关注系统的觉知、全局工作空间理论[41]、动机[700]、元认知、注意、内省学习等自动控制的问题。

### 8. 高级认知功能

脑的高级认知功能包括学习、记忆、语言、思维、决策、情感等。学习是通过神经系统不断接受刺激，获得新的行为、习惯和积累经验的过程，而记忆是指学习得到的行为与知识的保持和再现，是我们每个人每天都在进行着的一种智力活动。语言和高级思维是人区别于其他动物的最主要因素。决策是指通过分析、比较，在若干种可供选择的方案中选定最优方案的过程，也可能是对不确定条件下发生的偶发事件所做的处理决定。情感是人对客观事物是否满足自己的需要而产生的态度体验。

### 9. 动作选择

动作选择是指由原子动作构建复杂组合动作，以实现特定任务的过程。动作选择可以分为两个步骤，首先是原子动作选择，即从动作库选择相关的原子操作。然后，使用规划策略，将选定的原子动作组成复杂动作。动作选择机制可以基于尖峰基底神经节模型实现。

### 10. 响应输出

响应输出从总体目标开始运动分级，受外周区域输入的情感和动机的影响。基于控制信号，初级运动皮层运动区直接生成肌肉的运动，实现某种内部给定的运动命令。

关于 CAM 心智模型的详细讨论，请参阅著作 *Mind Computation*[692]。

## 4.9 PMJ 心智模型

认知心理学和认知神经科学的研究成果为阐明人类认知机理提供了大量的实验证据和理论观点,明确认知过程的主要阶段和通路,即感知、记忆和判断阶段以及快速加工通路、精细加工通路和反馈加工通路,傅小兰等构建了 PMJ(Perception,Memory,and Judgment,PMJ)心智模型[246],如图 4.10 所示。图中的虚线框内为心智模型,概括了认知的主要过程,包括感知、记忆和判断 3 个阶段(用齿轮圆表示)与快速加工、精细加工和反馈加工 3 类通路(用数字标注的带箭头的线表示)。在每个阶段,认知系统在各种认知机制的约束下,接受其他阶段的信息输入,完成特定的信息加工任务,并将加工结果信息输出到其他阶段。各个阶段相互配合,实现完整的信息加工过程。每类加工通路表示加工信息的传递。模型中还给出了认知与计算的对应关系,即感知阶段对应于计算流程中的分析,记忆阶段对应于计算流程中的建模,判断阶段对应于计算流程中的决策。

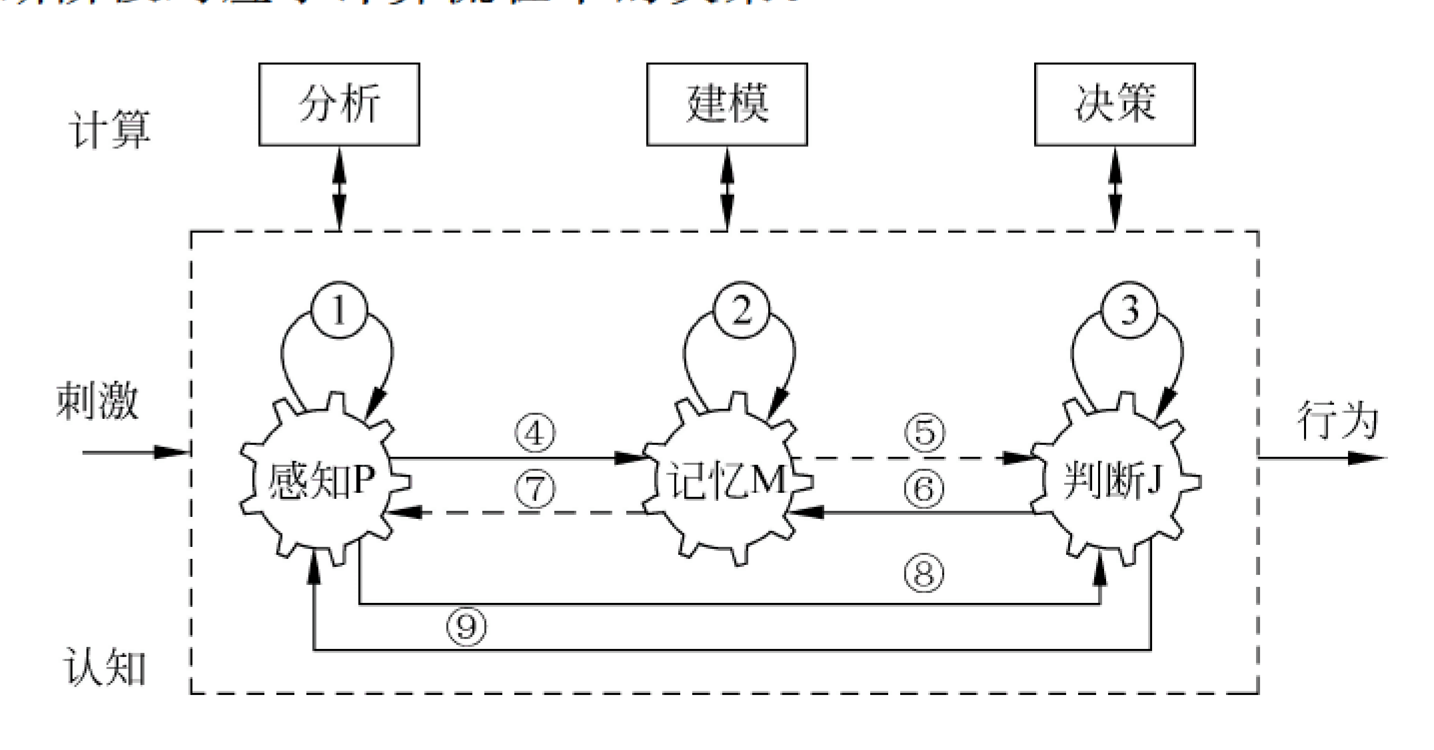



图 4.10 感知、记忆和判断心智模型

根据已有研究成果,PMJ 模型将认知加工的主要通路归纳为快速加工通路(类比于大细胞通路及其关联的皮层通路)、精细加工通路(类比于小细胞通路及其关联的皮层通路)以及反馈通反馈通路(自上而下的反馈)。

### 1. 快速加工通路

快速加工通路是指从感知阶段直接到判断阶段的加工过程(如图 4.10 中的⑧),实现基于感知的判断。该过程不需要过多的已有知识经验的参与,主要加工处理刺激输入的整体特征、轮廓以及低空间频率信息,对这些输入信息进行初级粗糙加工,在此基础上进行快速分类判断。视觉显著性特征可以通过快速加工通路进行分类判断。

### 2. 精细加工通路

精细加工通路是指从感知阶段到记忆阶段,再从记忆阶段到感知和判断阶段的加工过程(如图 4.10 中的④、⑤和⑦),实现基于记忆的感知和判断。该过程依赖于已有的知识经验,主要加工处理刺激输入的局部特征、细节信息以及高空间频率信息,并与长时记忆中存储的知识进行精细匹配,在此基础上进行分类判断。人们对外界的感知通常离不开注意,需

要通过注意从众多信息中将有用的信息筛检过滤，储存到记忆系统，继而形成表征。记忆表征自适应的动态记忆系统记忆空间内存储的信息也会随着认知加工活动的进行而动态地建构和变化。

### 3. 反馈加工通路

反馈加工通路是指从判断阶段到记忆阶段，或者从判断阶段到感知阶段（如图 4.10 中的⑥或⑨）的加工过程，实现基于判断的感知和记忆。认知系统根据判断阶段输出的结果，修正短时或长时记忆中存储的知识；判断阶段输出的结果也会给感知阶段提供线索，使感知阶段的信息加工更加准确高效。

## 4.10　动力系统理论

随着动力系统理论的研究，认知科学的动力系统基础理论在逐步形成，例如，格罗布斯[267]、罗伯特森(S S Robertson)[626]、西伦(E Thelen)和斯密斯(L B Smith)[755]的文章和著作给出了认知的动态理解思路。特别是冯・盖尔德(T van Gelder)和波特(R Port)于 1995 年出版了一本关于认知科学的动力理论的书[778]，提出认知科学的动力学研究思路，被作为认知科学第三种竞争范式的宣言。此书引起了较大凡响，如华盛顿大学伊莱斯密斯(C Eliasmith)1996 年发表了《第三种竞争范式：对认知的动力理论的批判性考察》[208]，其后也有其他人的热烈讨论。

冯・盖尔德针对 20 世纪 80 年代以后符号主义、联结主义范式所产生的困难，提出他的动力学假说。对于认知科学中的时间、构架、计算和表征等概念都提出了不同的解释。冯・盖尔德认为纽厄尔、西蒙的计算主义假说或物理符号系统假说：自然的认知系统在物理符号系统的意义上是智能的。

用动态眼光理解认知的还有丘奇兰德(P S Churchland)和谢诺沃斯基(Sejnowski)，他们把所拥护的联结主义假说表述为“突现性是以系统的某种方式依赖于低层现象的高层结果”[147]。他们认为通过构架的低层神经网络的作用能达到复杂的认知效果，直觉过程是一种亚概念的联结主义动力系统，它不接受完全的、形式化的、精确的概念层次的描述。用亚概念网络把自然认知系统看作是动力神经系统最好的理解。有一种假设认为，人意向性意识涌现于集群系统动力学，并由环境激发。

动力系统类包括任何随时间变化的系统，广泛用于对自然界的描述。动力论者期望勾画一类特殊的能恰当描述认知的动力系统。于是 1995 年冯・盖尔德给出他的动力学假说，认为自然的认知系统是某种动力系统，而且从动力学眼光理解认知系统是最好的理解。动力学假说是以数学的动力系统理论为基础描述认知的，用数学中的状态空间、吸引子、轨迹、确定性混沌等概念来解释与环境相互作用的智能体的内在认知过程。用微分方程组来表达处在状态空间的智能体的认知轨迹。换句话说，认知是作为智能体所有可能的思想和行为构成的多维空间被描述的，特别是通过在一定环境下和一定的内部压力下的认知主体的思想轨迹来详尽考察认知的。认知主体的思想和行为都受微分方程的支配。系统中的变量是不断进化的，系统服从于非线性微分方程，一般来讲是复杂的，是确定的。

下面简单介绍几种动力系统模型。

### 1. 循环动作行为模型(Cyclical Motor Behavior Model)

罗伯特森曾用动力学方法对新生婴儿的自发的动作行为中的循环做了大致勾画。罗伯特森采集了大量的关于新生婴儿呈现的自发的动作行为的数据。由于这些经验数据的有效性,这个动力系统模型是少有的几个能够充当动力系统模型的。而且许多人认为,这是一种可定量化的生理学行为的一种非隐喻的动力描述,恐怕较临床心理学的研究结果更能让人欣然接受。

罗伯特森后来过滤了观察状态空间,获得了带有少数自由度的一个理想的动力模型,似乎能够模拟循环动作的随机过程。但是至今还没有完美的动力系统模型。距离建立一种使状态变量和参数与生理学和环境因素有清楚对应的关于循环动作的动力系统模型的目标,还有相当长一段距离。

### 2. 嗅球模型(Olfactory Bulb Model)

1987年斯卡德(C A Skarde)和弗里曼(W J Freeman)的论文"为了了解世界大脑是如何制造混沌的"大致勾勒了这个模型并进行了一定程度的实验[725],这是一个基于嗅觉的神经过程的考察,借助复杂动力系统理论描述感受器官的神经系统的各种复杂状态,包括描述混沌神经元活动及其有规律的轨迹而提出的精致模型。在2000年,弗里曼提出了"介观脑动力学",由神经元到脑之间建立桥梁。冯·盖尔德、格罗布斯、巴顿(Bardon)和纽曼等都承认它可以作为动力系统模型。

### 3. 动力振荡理论模型(Motivational Oscillatory Theory)

动力振荡理论是一个关于循环的动力系统的模型,是冯·盖尔德推荐作为动力论假说范例的一个简化的动力系统模型。但是这个系统最大的问题就是如何正确选择系统的参数。因为对于动力系统而言,是对初值敏感的,"改变动力系统的一个参数就改变了它的整个动力学"[778]。

动力论的认知范式与其他范式的一个重要区别是对表示的不同理解。符号主义模型是以符号表示为基础的。联结主义的表示是以网络中的并行式表示或局部符号表示为基础的。但动力论的认知范式则宣称,一个动力模型应当是"无表示的"。

动力系统理论对认知行为的连续性提供了随时间变化的自然主义的说明。这是其他范式不能说明的,其他范式一般来讲是忽略时间概念的。但人类大脑与环境之间是随时有信息交流的,而且是不断变化的,暂态的连续的认知是随时间变化的。

动力系统理论的优势是对认知的描述是多元的,是一种经验可检验的理论,可以对描述认知系统的微分方程进行分析修正,也可以用已知的技术去求解这些方程,比起其他理论,它是一种定量的分析,是理解认知的一种确定性的观点。另一优势是动力系统的描述可以展示人类行为复杂、混沌的特性。动力论者认为,如此对认知的分析描述,应当是已经找到了替代认知科学中的符号主义、联结主义的新范式。如何保证动力系统的各变量和参数的恰当选择、系统的稳定性和可靠性问题,以及对于表示的理解等,是动力系统理论受到质疑之处。

## 4.11 大脑协同学

大脑中协同作用的科学研究，最早可以追溯到美国科学家斯佩里发现“裂脑人”不能实现左右脑的合作。以后斯佩里、康德尔和尚格等提出思维的神经回路理论。由于大脑有一千亿个神经元，每个神经元与三万个神经元相联系，形成一百万亿到一千万亿个接触点，因此形成大量的不同类型的神经回路。他们提出不同的回路与不同的思维相关。在著名的大脑功能定位学说中，研究神经元的结构和功能及神经网络的形成，一些神经元只感知个别信息，只有经过复杂的神经元的综合与协同，才能形成知觉。在思维生理学中，研究大脑活动机制，其反映为脑电波，由脑电图、脑磁图的特点和变化可以了解思维。苏联科学家鲁比亚在《神经生理学原理》中提出大脑的 3 个不同功能联合区，它们彼此协作。

哈肯（H Haken）是协同学的创立者，协同学（即“协同工作之学”——哈肯语）是系统科学和非线性科学的基础理论之一。它把耗散视为自组织的条件，把协同当作自组织的动力，从一个崭新的角度揭示了非平衡态中自组织的形成和发展过程的规律。哈肯特别专注于协同学在脑科学和人工智能等学科中的应用研究，先后发表了《协同计算机和认知——神经网络的自上而下方法》[298]和《大脑工作原理——脑活动、行为和认知的协同学研究》[299]两部最有代表性的专著。前者根据“协同形成结构，竞争促进发展”这一相变过程中的普遍规律，提出了“协同计算机”和“协同神经网络”的新概念，指出模式识别就是模式形成，并描述了自上而下的协同计算机构造方法。后者更直接地将非平衡自组织理论运用于人脑这一最复杂系统机理的探究，提出大脑是一种具有涌现性的复杂自组织巨系统的新见解，并建立了用以详尽阐述以上新见解的大量实验结果的具体模型。正如哈肯教授所说“这些模型皆用一个统一观点——协同学观点——加以表述”，我们不妨称之为“大脑协同学认知模型”。哈肯建立的协同学认知模型，运用协同学的一般原理和方法，提出了大脑工作的新见解——大脑是一种具有涌现性的复杂自组织巨系统，从而对大脑功能做出了协同学的解释。

复杂性是开放的复杂巨系统的一种很重要的特性，研究复杂性离不开系统。Santa Fe（圣塔菲）研究所的代表性工作是提炼出了许多很有意义的概念，比如人工生命与混沌的边缘，基因网络与自催化系统，自组织的临界性，复杂自适应系统等等。这些概念共同构成了这个研究机构的哲学基础，它以后的工作都是在这一框架下展开的。圣塔菲的研究并没有对生命起源、涌现等问题做出令人满意的答复，但是它所提出的每一个新概念都代表着一种新的态度、一种看待问题的新角度和一种全新的世界观。

加拿大籍奥地利理论生物学家贝塔朗菲（Luduig von Bertalanffy）开创了现代的系统科学。他提出一般系统论是为了阐明对于有生命的物体来说，“整体大于部分之和”。也就是说，系统的特征是不能由孤立的各部分的特征来说明的，因此复合体的特征与元素的特征相比是“新的”或“突然发生的”。他反对生物学中机械论的思想，强调生物学中有机体概念，主张把有机体当作一个整体或系统来考虑，认为生物学的主要任务应当是发现生物系统中一切层次上的组织原理。贝塔朗菲认为机械论的观点是错误的，其主要错误观点：一是简单相加的观点，即把有机体分解为各要素，并采用简单地相加来说明有机体的属性；二是机械

观点,即把生命现象简单地比作机器;三是被动反映的观点,即把有机体看作只有受到刺激时才能反映,否则就静止不动。他概括地吸取了生物机体论的思想,并加以发展,提出了新的机体论思想,其主要观点:一是系统观点,认为有机体都是一个系统,并把系统定义为相互作用的诸要素的复合体;二是动态观点,认为一切生命现象本身都处于积极的活动状态,活的东西的基本特征是组织:主张从生物体和环境的相互作用中说明生命的本质,并把生命机体看成是一个能保持动态稳定的系统;三是分层观念,认为各种有机体都是按严格的层级组织起来的,生物系统是分层次的,从活的分子到多细胞个体,再到超个体的聚合体,可谓层次分明,等级森严。

普里高津(Ilya Prigogine)对自组织的研究,以及提出所谓的耗散结构理论是对新的东西如何呈现出来的机理的进一步探讨。在他与尼柯利斯(G Nicolis)合著的《探索复杂性》中他们表达了自己的指导思想:他们所反叛的是传统物理学家对世界的经典认识观点。自从牛顿以来,可逆性与决定性是物理学家继续经典研究项目的传统理念[541]。但是,无数的科学发现使得人们认识到发生在自然界中的许许多多的基本过程是不可逆的、随机的,那些描述基本相互作用的决定性和可逆性的定律不可能告诉人们自然界的全部真相。而且研究发现在远离平衡态的情况下,分子之间可以互相传递信息,这样对处于远离平衡态的世界进行研究,就可以跨越自然科学的范围而进入人文科学的领域。而相互通信这一点就是维纳在构造他的理论体系时所用的基本概念之一,通过互传信息实现了控制的产生。基于这些理解和认识,普里高津和尼柯利斯将非线性非平衡态系统的概率分析方法同动力学理论,特别是混沌动力学理论所表达的决定性的系统也可以与初始条件很敏感这一特性相结合,从而解释了在我们所处的环境中还有如此多意想不到的规律性。

哈肯于1971年提出协同的概念,1976年系统地论述了协同理论,发表了《协同学导论》,进一步发展了普里高津对这个问题的研究。他们所考虑都是远离平衡态的相变,但是这种从微观或中观到宏观的转变都是有条件的。协同论认为,千差万别的系统,尽管其属性不同,但在整个环境中,各个系统间存在着相互影响而又相互合作的关系。其中也包括通常的社会现象,如不同单位间的相互配合与协作,部门间关系的协调,企业间相互竞争的作用,以及系统中的相互干扰和制约等。协同论指出,大量子系统组成的系统,在一定条件下,由于子系统相互作用和协作,这种系统的研究内容,可以概括地认为是研究从自然界到人类社会各种系统的发展演变,探讨其转变所遵守的共同规律。应用协同论方法,可以把已经取得的研究成果,类比拓宽于其他学科,为探索未知领域提供有效的手段,还可以用于找出影响系统变化的控制因素,进而发挥系统内子系统间的协同作用。

哈肯认为宏观是指空间、时间或者功能结构,而这些结构对比于所考虑的每一个微观或者中观粒子的性质来说,只不过是一种累加行为而已,是在概率意义上的累加。对于一个描述动力系统的非线性微分方程组来说,采用线性方法进行稳定性分析得出不稳定结果时,在某些条件下可能通过变换变量或方程的方法将变量和方程组的个数缩减为很少几个,对原动力系统的定性分析完全可以通过分析经过缩减后的方程组得到。哈肯在1996年的《大脑工作原理》中,系统阐述了他的脑活动和认知的协同学研究结果。大脑功能的传统实验和理论研究以单个细胞为依据,而协同学的注意力集中在整个细胞网络的活动上。表4.2给出了它们对有关术语的不同解释。

表 4.2　大脑功能的传统解释与协同学解释

| 传统解释 | 协同学解释 |
| --- | --- |
| 细胞 | 细胞网络 |
| 个体 | 整体 |
| 祖母细胞 | 细胞集体 |
| 引导细胞 | 细胞集体 |
| 定域的 | 非定域的 |
| 兴奋印迹 | 分布信息 |
| 编程计算机 | 自组织的 |
| 算法的 | 自组织的 |
| 序贯的 | 并行和序贯的 |
| 确定性的 | 确定性事件和偶然事件 |
| 稳定的 | 趋于不稳定点 |

表 4.2 概括了哈肯研究所取得的一些基本结果，也是理解其“大脑协同学”理论的关键点。简言之，“大脑是遵从协同学规律的复杂巨系统，即系统运转在趋于不稳定点处，由序参量决定宏观模式”，即通过各个部分的相互作用，系统以自组织方式在宏观层次上涌现出全新的属性。这种属性在微观层次的各个细胞中是不存在的。正因如此，哈肯才说：“虽然神经计算机的发展在模拟神经元活动方面确实迈出了非常重要的一步，但我相信，以一般协同学概念为基础的协同计算机，更接近认识脑活动这一目标。”据说，协同计算机的理论设计和模式识别效果，都比神经计算机先进许多。他由此提出了协同计算机的三层网络模型，并强调不应把认知系统看作代表外部环境的内部网络，而应当看作内部-外部网络。同时又指出现代计算机距离能够真正思考还很遥远，而脑研究可为我们提供目前意想不到的洞见，主张人工智能与脑科学之间的协作，这正好印证了其“协同学”的第二重含义：“完全不同的学科之间的协作、碰撞，进而产生一些新的科学思想和概念。”

哈肯曾经预言，从长远的观点看，有希望制造出以自组织方式执行程序的协同计算机来模拟人类智能。

## >>第 5 章

# 视觉感知

视觉系统使生物体具有视觉感知能力。它使用可见光信息构筑机体对周围世界的感知。根据图像发现周围景物中有什么物体和物体在什么地方的过程,也就是从图像得到对观察者有用的符号描述的过程。视觉系统具有将外部世界的二维投射重构为三维世界的能力。需要注意的是,不同物体所能感知的可见光处于光谱中的不同位置。

视觉主要有两个功能:一是目标知觉,即它是什么?二是空间知觉,即它在哪里?已有确实的证据表明,不同的大脑系统分别参与上述两种功能。

## 5.1 视觉的生理机制

眼是视觉的外周器官,是人体所有感官中最复杂的。来自外界的信息很大部分是通过眼接收的。外界物体的光线射入眼中,聚焦后在视网膜上成象。从视网膜发出神经冲动,经视神经达到大脑皮层视区,产生视觉。

视觉在人类的感觉世界中担负着重要的任务。我们对大部分环境信息作出反应,是经过视觉传入大脑的。它在人类的感觉系统中占主导地位。如果人类用视觉接收一个信息,而另外一个信息是通过另一个感觉器官接收的,又如果这两个信息相互矛盾,人们所反应的一定是视觉信息。

### 5.1.1 视网膜

光线进入眼到达视网膜主要经过三个折光面:①空气——角膜界面;②房水——晶状体界面;③晶状体——玻璃体界面。眼的折光系统由四种折射率不同的介质组成,又有几个曲率不同的折射面。要准确地描述光线在眼球内的折射路径是可能,但太复杂。通常采用简化的方法,即简约眼,也可以足够准确地求得光线通过眼的路径。

视网膜是脑的一部分,它是由处理视觉信息的几种类型的神经元组成的。它紧贴在眼球的后壁上,厚度只有 0.5mm 左右。包括三级神经元:第一级是光感受器,由无数视杆细胞和视锥细胞组成;第二级是双极细胞;第三级是神经节细胞。由神经节细胞发出的轴突形成视神经。这三级神经元构成了视网膜内视觉信息传递的直接通道。此外,在第一级与第二级以及第二级与第三级神经元之间,还有两种中间神经元:水平细胞和无长突细胞。它们的突起在细胞层之间水平延伸,把相邻的神经元联系起来。图 5.1 给出了灵长类动物的视网膜结构[967]。

视杆和视锥细胞是眼球的光受纳器细胞。视杆细胞细而长,视锥细胞粗而短。它们对光敏感性方面有明显的差别。视杆细胞能感受极其微弱的光线,但不能分辨颜色;视锥细

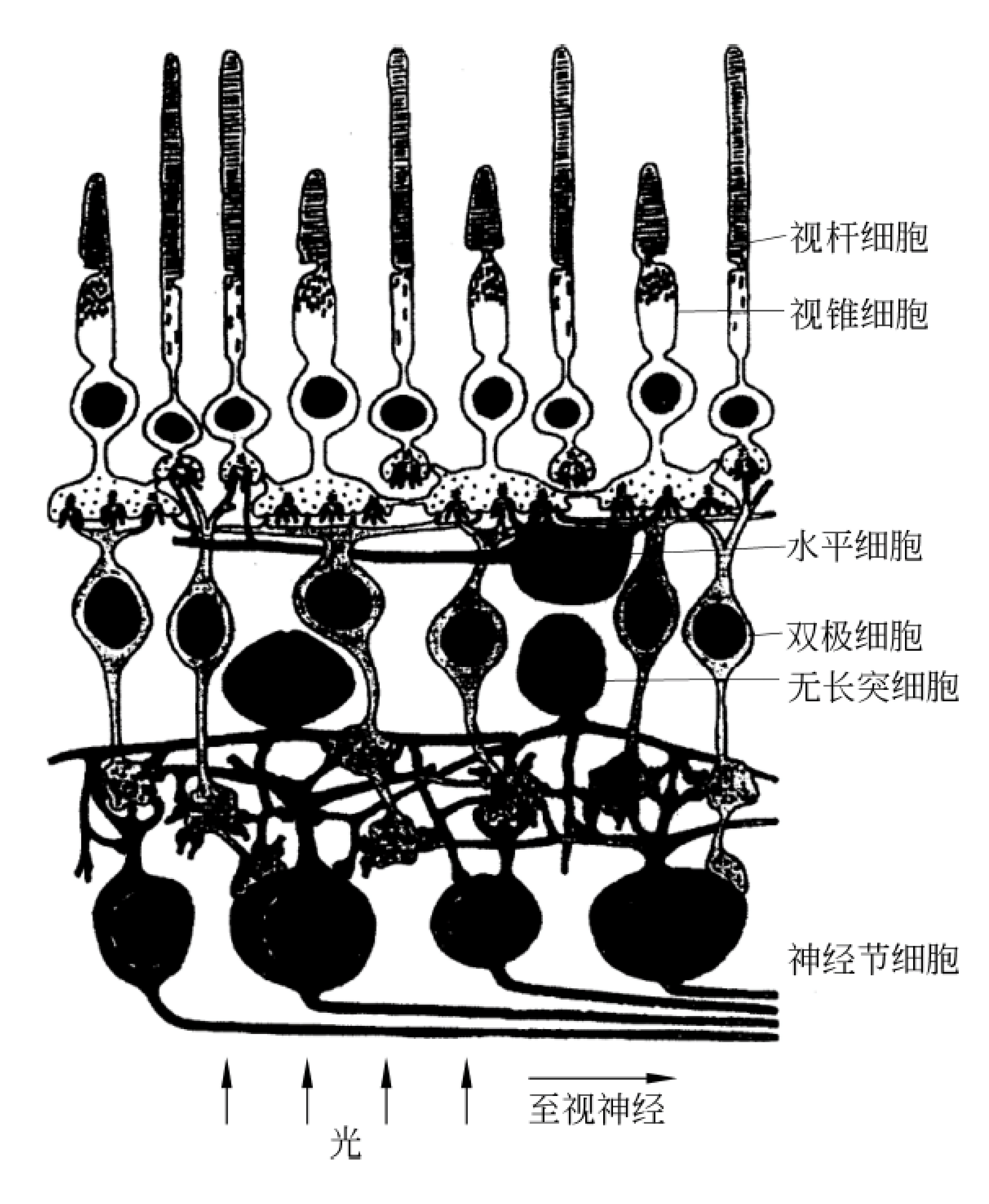



图 5.1 灵长类动物的视网膜结构

胞只对较强的光线产生反应，能感受不同的颜色。视网膜上视杆细胞的数目比视锥细胞要多得多。人的视网膜上共有 600 万～700 万个视锥细胞，然而视杆细胞的数目高达 1.1～1.3 亿个。与如此庞大数量的感受细胞相比，视网膜的传出神经元——神经节细胞的数目大约只是其百分之一。有数十个～数百个感受细胞通过双极细胞(以及水平细胞和无长突细胞)与一个神经节细胞相联系。在视网膜的中心区域，叫做视斑，在这里，锥体细胞占绝大多数。在视斑的中心，叫做中央凹，在这里全部是锥体细胞。中央凹的面积只有 $1mm^2$，但是它产生最清晰的视觉，因为晶体将物象聚焦在中央凹上，也因为在这个区域与高级脑中枢有一对一的连接在视网膜中央凹范围内，视锥细胞的密度最高。视杆细胞的密度在中央凹处很低，而在旁中央区最高。神经节细胞的分布大致与视锥细胞的相同。这种分布特点决定了中央区视野具有极高的视敏度(分辨空间细节的能力)，而旁中央区视野对弱光最敏感。

## 5.1.2 光感受器

人类的视网膜内有四种光感受器：视杆细胞和三种视锥细胞。在每一种感受器内都含有一种特殊的色素。当一个这样的色素分子吸收了一个光量子以后，它会在细胞内触发一系列的化学变化；与此同时释放出能量，导致电信号的产生和突触化学递质的分泌。视杆细胞的视色素称“视紫红质”，其光谱吸收曲线的峰值波长为 500nm。三种视锥细胞色素的光谱吸收峰值分别在 430nm、530nm 和 560nm，分别对蓝、绿、红三种颜色最敏感。

感受器与双极细胞之间突触的性质(兴奋或抑制)可能决定于感受器所释放的递质种类，也可能决定于双极细胞突触后膜的通道性质。目前还没有证据表明同一个感受器能够释放两种递质。更多的证据显示这两种双极细胞可能有不同的受体分子。双极细胞感受野

的中心区是通过感受器与双极细胞的直接联系形成的,其外周区则是通过水平细胞与感受器的间接联系形成的。

每一个感受器同时既与双极细胞也与水平细胞建立联系。水平细胞通过其树突的分枝在水平方向上联系附近的许多感受细胞,然后将总合的信号传递给双极细胞。感受器与水平细胞之间的信息传递是通过兴奋性突触。由于水平细胞没有轴突,目前还不清楚灵长类动物水平细胞的输出端究竟在哪里。通过水平细胞的横向联系,形成了双极细胞和神经节细胞的感受野外周区。水平细胞的感受野是一种均匀的结构,光刺激任何位置都引起均一的超极化反应。刺激光点越大,超极化也越强。

无长突细胞的胞体在双极细胞和神经节细胞之间,通过树突的广泛分支在水平方向上联系附近的许多神经节细胞。它们没有轴突,其树突兼有突触前和突触后的作用。无长突细胞彼此之间也有相互联系。由于这个特点,这种细胞空间总和的范围比水平细胞要大得多。无长突细胞的形态各式各样。使用的神经递质的种类也十分复杂。

神经节细胞是视网膜的输出级神经元。一个神经节细胞可以通过双极细胞和水平细胞联系许多个感光细胞的输入,而相邻的神经节细胞之间,又可以通过无长突细胞的水平分枝形成相互联系。神经节细胞与其他视网膜细胞之间的这些复杂神经联系最终反映在它们的感受野结构上。与所有前级神经元不同,从神经节细胞开始才产生了传导的动作电位。可以通过记录神经节细胞或视神经的电脉冲来研究该细胞的功能特性和感受野结构。神经节细胞的感受野和双极细胞的一样,包括一个中心区和一个外周区。无论是哪一类细胞,感受野中心区与周边区之间在功能上都是相互拮抗的,光刺激周边区会减弱对中心区刺激的反应强度。

### 5.1.3 外膝体

视神经在进入脑中枢前以一种特殊的方式形成交叉。从两眼鼻侧视网膜发出的纤维交叉到对侧大脑半球;从颞侧视网膜发出的纤维不交叉,投射到同侧大脑半球。其结果是:从左眼颞侧视网膜来的纤维和从右眼鼻侧来的纤维汇聚成左侧视束,投射到左侧外膝体;再由左外膝体投射到左侧大脑半球,与相应脑区对应的是右侧半个视野。相反,从左眼鼻侧视网膜来的纤维和从右眼颞侧视网膜来的纤维汇聚成右侧视束,投射到右侧外膝体;再由右侧外膝体投射到右侧半球,相应脑区对应于左侧半个视野。脑两个半球的视皮层通过胼胝体的纤维互相连接。这种相互联结,使从视野两边得来的信息混合起来。

视觉信息在视网膜内初步加工后,通过视神经和视束纤维传入丘脑。在丘脑内分成两条通路:①外膝体-皮层通路,是主要的视觉传入通路。在灵长类,从视网膜发出的神经纤维中有 90%经由这条通路投射到视皮层。②上叠体-丘脑枕-皮层通路,灵长类只有 10%网膜纤维经由这条通路与视皮层建立联系。在丘脑枕内与视觉有关的主要是腹外侧丘脑枕和下丘脑枕,分别接受来自视网膜和上叠体的传入纤维。这一条通路的神经联系和功能特性还有待研究。

灵长类的外膝体由六层细胞构成。两眼的传入分别投射到三个不同层次。从同侧眼(相对于外膝体)来的纤维终止于第 5、3、2 层;从对侧眼来的纤维终止于 6、4、1 层。这些相互重叠的每一个层次都与视网膜有点对点的投射关系。因此在视网膜上相邻近的点,在外膝体的每一层上也是相邻的。由于这 6 层是按网膜投射整齐地重叠起来的,因此,如果在第

6 层的某一点的细胞从左侧视网膜的一个特定点接受输入，那么从第 5 层的相应点必定也是从右眼的该对应点接受输入。这种解剖上的对应关系有利于建立双眼间的相互联系，为在视皮层形成双眼视差（深度和立体感）调谐奠定了基础。

在视网膜和外膝体神经元的传统感受野外，还存在着一个范围相当大的区域。这个区域对视觉刺激不产生直接反应，但对感受野内刺激所引起的反应有调制作用。对一个给光中心细胞来说，当刺激光点超出感受野中心的大小时会引起视觉神经元反应的减小，这表明在感受野中心兴奋区的外面，有一个抑制性的外周区。去抑制区与感受野之间的颜色相互作用可能是产生颜色对比和色觉恒定性等心理现象的神经基础。色觉恒定性和颜色对比是人色觉的两个重要特性。所谓色觉恒定性，是指视觉系统可以在周围环境光谱成分出现明显变化的情况下保持对物体颜色的认知相对不变。颜色对比是指在绿色背景下，人眼对红色的敏感性增强（或相反）。当周围环境光线趋红时，通过去抑制区的作用，外膝体神经元能自动地降低对红光的敏感性，提高对蓝光的敏感性，反之亦然。从而保持了颜色知觉不变，并增强了颜色对比。

### 5.1.4　视皮层

视皮层本身的神经元主要有两种：星形细胞和锥体细胞。星形细胞的轴突与投射纤维形成联系。锥体细胞呈三角形，尖端朝表层，向上发出一个长的树突，基底则发生几个树突作横向联系。

视皮层和其他皮层区一样，包括 6 个细胞层次，由表及里用罗马数字 Ⅰ～Ⅵ来代表。皮层神经元的突起（树突和轴突）的主干都沿与皮层表面相垂直的方向分布；树突和轴突的分枝则横向分布在不同层次内。不同皮层区之间由轴突通过深部的白质进行联系，同一皮层区内由树突或输突在皮层内的横向分枝来联系。

近年来，视皮层的范围已扩大到顶叶、颞叶和部分额叶在内的许多新皮层区[776]，总数达 25 个。另外还有 7 个视觉联合区，这些皮层区兼有视觉和其他感觉或运动功能。所有视区加在一起占大脑新皮层总面积 55%。由此可见视觉信息处理在整个脑功能中所占有的分量。研究各个视区的功能分工、等级关系以及它们之间的相互作用，是当前视觉研究的一个前沿课题。确定一个独立的视皮层区的依据是：①有独立的视野投射图，该区与其他皮层区之间有相同的输入和输出神经联系；②该区域内有相似的细胞筑构；③有不同于其他视区的功能特性。

韦尼克（Wernicke）和格什温德（Geschwind）认为，视觉识别的神经通路如图 5.2 所示。根据他们的模型，视觉信息由视网膜传至外侧膝状体，从外侧膝状体传至初级视皮层（17 区），然后传至一个更高级的视觉中枢（18 区），并由此传至角回，然后至 Wernicke 区[472]。在 Wernicke 区，视觉信息转化为该词的语声（听觉）表象。声音模式形成后，经弓状束传至 Broca 区。

视皮层中 17 区被称为第一视区（V1）或纹状皮层。它接受外膝体的直接输入，因此也称为初级视皮层。对视皮层的功能研究大多数是在这一级皮层进行的。除了接受外膝体直接投射的 17 区之外，和视觉有关的皮层还有纹前区（18 区）和纹外区（19 区）。根据形态和生理学的研究，17 区不投射到侧皮层而仅射到 18 区，18 区向前投射到 19 区，但又反馈到 17 区。18 区内包括三个视区，分别称为 V2、V3 和 V3A，它们的主要输入来自 V1。V1 和

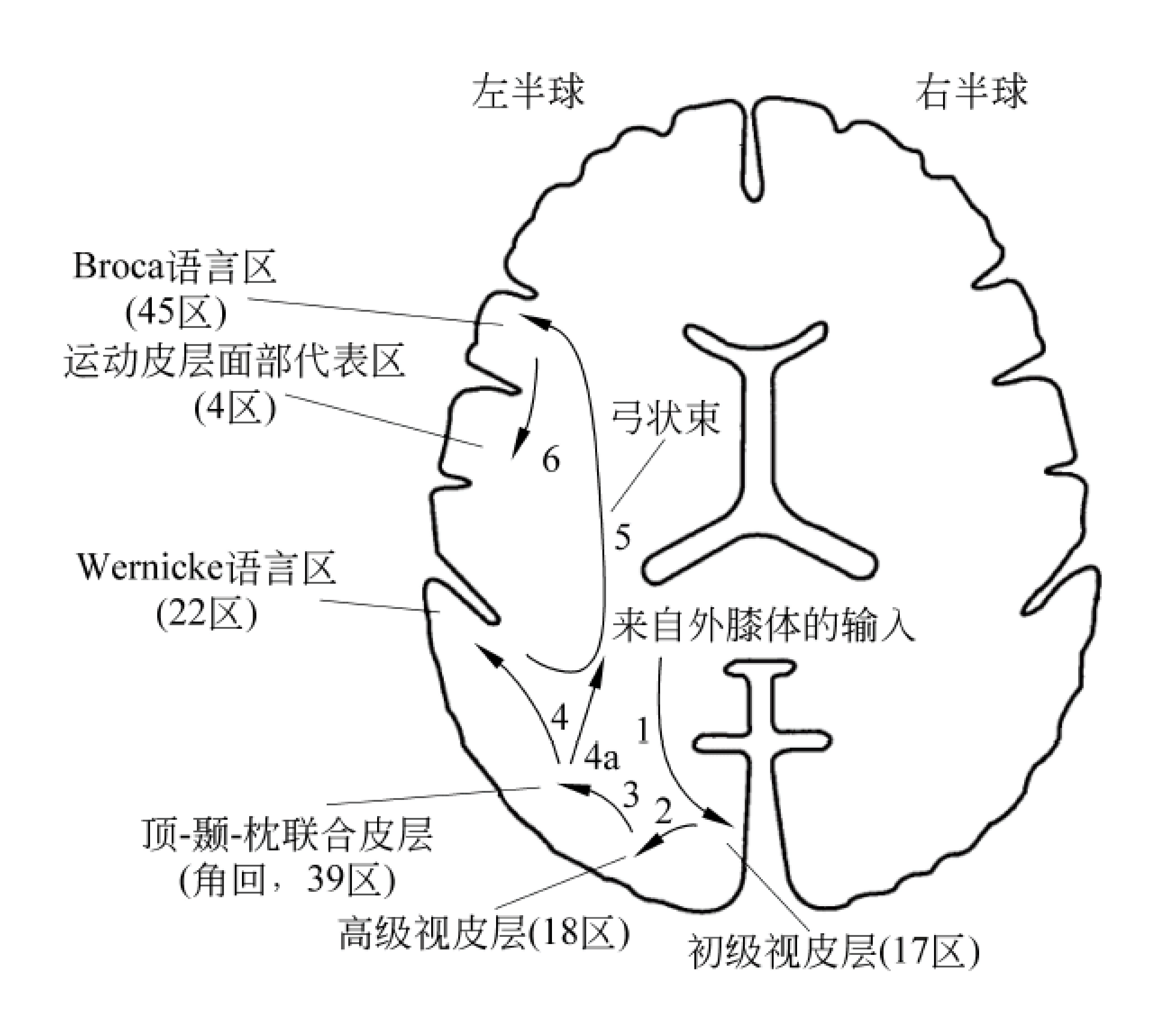



图 5.2 视觉的神经通路

V2 是面积最大的视区。19 区深埋在上颞沟后壁，包括第四(V4)和第五视区(V5)。V5 也称作中颞区，已进入颞叶范围。颞叶内其他与视觉有关的皮层区还有内上额区、下颞区。顶叶内有顶枕区、腹内顶区、腹后区和 7a 区。枕叶以外的皮层区可能属于更高的层次。为什么要这样多的代表区？是不是不同代表区检测图形的不同特征(如颜色、形状、亮度、运动、深度等)？或是不同代表区代表处理信息的不同等级？会不会有较高级的代表区把图形的分离特征整合起来，从而给出图形的生物学含义？是不是有专门的代表区负责储存图像(视觉学习记忆)或主管视觉注意？这些都将是在一个更长的时间内视觉研究有待解决的问题。

### 5.1.5 感受野

视皮层神经元对光点刺激的反应很弱，只有在感受野内用适当方位(朝向)的光条给以刺激才能引起兴奋。根据皮层神经元感受野结构的不同，休贝尔(Hubel)和维塞勒(Wiesel)对猫和猴的视皮质中单一神经元的激发模式进行的研究，发现有四种类型视皮层神经元——简单细胞、复杂细胞、超复杂细胞和极高度复杂细胞。

简单细胞感受野面积比较小，给光区和撤光区分离，有较明显的空间总合，反应具有线性特征，没有(或很少)自发放电。刺激具有特定的方向和在视野中有固定的位置，最能激发简单细胞。

复杂细胞感受野面积较大，光区和撤光区重叠在一起，反应具有非线性特征，空间总合不明显，自发放电强。与简单细胞不同之处是，复杂细胞反应要求一定方位的线性刺激，但不管在视野中的部位如何。另一个与简单细胞不同之处是，当光线移过视野时，能继续激发复杂细胞对于适当方位的线性刺激。简单细胞对刺激在视野中的位置改变时，则停止激发，刺激的方位保持不变。因为复杂细胞对于适当方位的移动的直线刺激能继续激发，可以认为它们接受从大量的简单细胞输入的刺激，所以这些简单细胞都反应某种方位的刺激，但每一简单细胞对视野中不同部位的刺激产生感觉。

超复杂细胞能从几种复杂细胞中接受兴奋性的和抑制性的输入信息。超复杂细胞也反

应特殊方位的线性刺激,但是,这种刺激不能超过某种长度。反应特点同复杂细胞,但有明显的端点抑制。

极高度超复杂细胞反应移过视野的边,只要是这边有一特定的宽度。有些极高度超复杂的神经元特别反应两个边形成的 90°的角,因此,这种极高度超复杂神经元,又称为角探察器。

视皮层神经元对视觉刺激的各种静态和动态特征都具有高度选择性,包括方位/方向选择性、空间频率选择性、速度选择性、双眼视差选择性、颜色选择性。

### 1. 方位/方向选择性

视皮层细胞只有当刺激线条或边缘处在适宜的方位角并按一定的方向移动时,才表现出最大兴奋(最佳方位或最佳方向)。以细胞的放电频率相对于刺激方位和运动方向作成直方图,可以显示该细胞的方位和方向调谐特性。

### 2. 空间频率选择性

正弦波调制的光栅是视觉实验中经常使用的刺激图形。用这种刺激图形的主要优点是便于对视觉反应的时空特性进行定量的数学分析。每一个视皮层细胞都有一定的空间频率调谐。在同一皮层区内,不同细胞也有不同的空间频率选择性。

### 3. 速度选择性

视皮层细胞对移动图形的反应比对静止的闪烁图形要强得多。每一个皮层细胞不仅对运动的方向有选择性,而且要求一定的运动速度。只有当刺激图形在适宜的方向上以某一定速度移动时,细胞反应才达到最大。这个速度称为该细胞的最佳速度。移动速度高于或低于最佳速度时,反应都会减小。

### 4. 双眼视差选择性

与外膝体细胞不同,大部分视皮层细胞接受双眼输入。因此,每一个细胞在左、右视网膜上都有一个感受野,这一对感受野在视网膜上的位置差(相对于注视点)称为“视差”(disparity)。如果左、右感受野与注视点的距离差为零,表示该点正好在注视平面上。如果两个感受野都向额侧偏离,表示该细胞的调谐距离(最佳距离)比注视点远;两个感受野向鼻侧偏离则意味着该细胞的调谐距离比注视点近。

### 5. 颜色选择性

同视网膜和外膝体神经元一样,皮层细胞也具有颜色选择性。与皮层下的单颉颃式感受野不同,视皮层细胞的颜色感受野具有双颉颃式结构。例如,对于 R-G(红-绿)型感受野来说,其颜色结构可能有两种形式。感受野中心可能被绿视锥细胞的输入兴奋,同时被红视锥细胞输入抑制,或者相反。外周对颜色的反应性质正好与中心相反。因此,双颉颃式感受野通过中心的颜色颉颃能分辨红色和绿色,通过中心与外周之间的相互作用能使红－绿对比的边缘得到增强。对于 B-Y(蓝-黄)型感受野,情况也一样。

### 5.1.6 功能柱

视皮层功能柱的发现是近年来视觉中枢研究中最引人注目的进展之一。由于这些柱系统正好与各种特征检测功能一一对应,所以也被认为是对特征检测理论的一个有力支持。到目前为止所发现的各种功能柱都垂直于皮层表面,排列成片层状。

#### 1. 方位柱

电生理学和形态学观察都已反复证明,在猴和猫的17区和18区存在着方位柱结构。Hubel和Wiesel用微电极倾斜地穿刺视皮层,在微电极通路上,依次被记录到的细胞的敏感方位总是很有规律地按顺时针或逆时针方向变化。方位柱是与皮层表面垂直排列的。

#### 2. 眼优势柱

休布尔(Hubel)和威塞尔(Wiesel)根据细胞受双眼影响程度的不同,将视皮层细胞分为左眼优势细胞与右眼优势细胞。他们在研究方位柱的同时还观察到,微电极垂直插入时所记录到的细胞串倾向于有相同的眼优势特性,然而在倾斜穿刺时,左眼优势细胞与右眼优势细胞通过一定的间隔交替出现。可见视皮层神经元的眼优势特性也是排列成垂直柱形式。每个柱的宽度大约500μm,左眼柱与右眼柱合在一起也是占1mm皮层范围。后来,用多种形态学方法(变性轴突法、还原银法和放射自显影法)都成功地显示了眼优势柱的结构。

#### 3. 空间频率柱

汤普森(Thompson)和托尔赫斯特(Tolhurst)用类似的方法观察到,在微电极穿刺中也出现细胞空间频率特性的系统变化,他们认为,皮层细胞的最佳空间频率也是有规则地以柱的形式垂直于皮层表面排列的。托尔赫斯特用2DG放射自显影方法证明,猫皮层17区确实存在着空间频率柱结构。这种柱系统垂直于皮层表面,从表层延伸到深层,柱的宽度为0.8～1mm。

#### 4. 颜色柱

Micheal在研究猴视皮层的颜色特异性时发现,有些细胞只对颜色光敏感,对白光没有反应。他还注意到,当微电极倾斜穿刺视皮层时,有颜色特异性的细胞和没有颜色特异性的细胞是成串地交替出现的。根据在几个平行穿刺通道上所得到的结果,他们认为,颜色特异性细胞也是排列成片层状结构。颜色柱宽度为100～200μm,贯穿整个皮层厚度,同一柱内所有细胞有相同的光谱特性。

中国科学院上海生命科学研究院神经科学研究所李朝义院士及其学生姚海姗共同研究发现,在初级视皮层中,存在着一种与处理大范围复杂图形特征有关的球状功能结构,新发现的球状结构是一种全新的脑功能结构,对了解脑处理复杂图像信息的神经机制有很重要的意义。电脑至今无法像人的视觉系统那样从千变万化的图像背景中迅速辨认出要搜寻的目标,而人脑却能轻易做到,传统的功能柱研究还不能阐释视觉系统究竟是如何处理大范围复杂图像信息的。此次两位中国科学家在猫的视皮层中,用一种特殊的微电极记录了几百个神经细胞,详细描绘了这些细胞的大范围整合特性以及对图形的反应,确定了在初级视皮

层中存在着一种与处理大范围复杂图形特征有关的功能结构。

### 5.1.7 颜色视觉

颜色是一个主观感觉。颜色可分为非彩色和彩色两大类。非彩色指白色到黑色及其间各种深浅不同的灰色,它们可以排列成一个系列,叫做黑白系列。彩色系指黑白系列以外的各种颜色,彩色有三种特性:明度、色调和饱和度。明度系指彩色光的亮度。色调是彩色彼此相互区分的特性,光源的色调决定于辐射的光谱组成对人眼所产生的感觉。色饱和度是指彩色的纯洁性。可见光谱中各种单色光是最饱和的彩色。当光谱色掺入白光成分越多,就越不饱和。

1854 年格拉斯曼(H Grassmann)建立了颜色混合定律。

(1) 人的视觉只能分辨颜色的三种变色:明度、色调和饱和度。

(2) 在由两个成分组成的混合色中,如果一个成分连续地变化,混合色的外观也连续地变化。由此定律又导出如下两个定律。

补色定律:每一种颜色都有一个相应的补色,它们以适当比例混合,产生白色和灰色;如果两者按其他比例混合,便产生近似比例大的颜色成分的非饱和色。

中间色定律:任何两个非补色相混合,其色调决定于两颜色的相对数量,其饱和度决定于二者在色调顺序上的远近。

(3) 凡视觉上相同的颜色,不管其光谱组成是否一样,在颜色混合中都是等效的。由此定律又导出颜色代替定律。

代替定律:相似色混合后仍相似。

(4) 混合色的总亮度等于组成混合色的各种颜色光亮度的总和,即亮度相加定律。

颜色视觉过程可以分成几个阶段。第一阶段,视网膜有三种不同的锥体细胞,它们各有独立的视色素能够选择地吸收不同波长的可见光,同时每一物质又可单独产生白-黑反应,在强光作用下产生白的反应,无光刺激时是黑色的反应。第二阶段,在神经兴奋由锥细胞向视中枢传递过程中,这三种反应又重新组合,最后形成三对颉颃的神经反应,即红-绿、黄-蓝,白-黑反应。这一过程的示意图见图 5.3。现在认为,扬格-黑尔姆霍兹(Young-Helmholtz)三色学说和赫林(Hering)的拮抗学说都只是对颜色视觉的一个方面获得了正确的认识,而必须通过两者的相互补充才能对色觉获得较为全面的认识。1807 年,扬格根据红、绿、蓝三原色可以产生各种色调和灰色的颜色混合规律,提出假设视网膜上有三种神经纤维,每种纤维的兴奋都引起一种原色的感觉,即红、绿、蓝色觉。其后,黑尔姆霍兹 1860 年进一步对三色说作了补充,提出光谱的不同成分引起三种纤维不同比例的兴奋,它们有不同的兴奋曲线的假说。1878 年,赫林观察到颜色现象总是以红-绿、黄-蓝、黑-白成对关系发生的,因此他假定视网膜中有三对视素:白-黑视素、红-绿视素、黄-蓝视素。每对视素对光照和黑暗呈相反的反应。

### 5.1.8 知觉恒常性

知觉恒常性是指人能在一定范围内不随知觉条件的改变而保持对客观事物相对稳定特性的组织加工的过程。它是人们知觉客观事物的一个重要的特性。

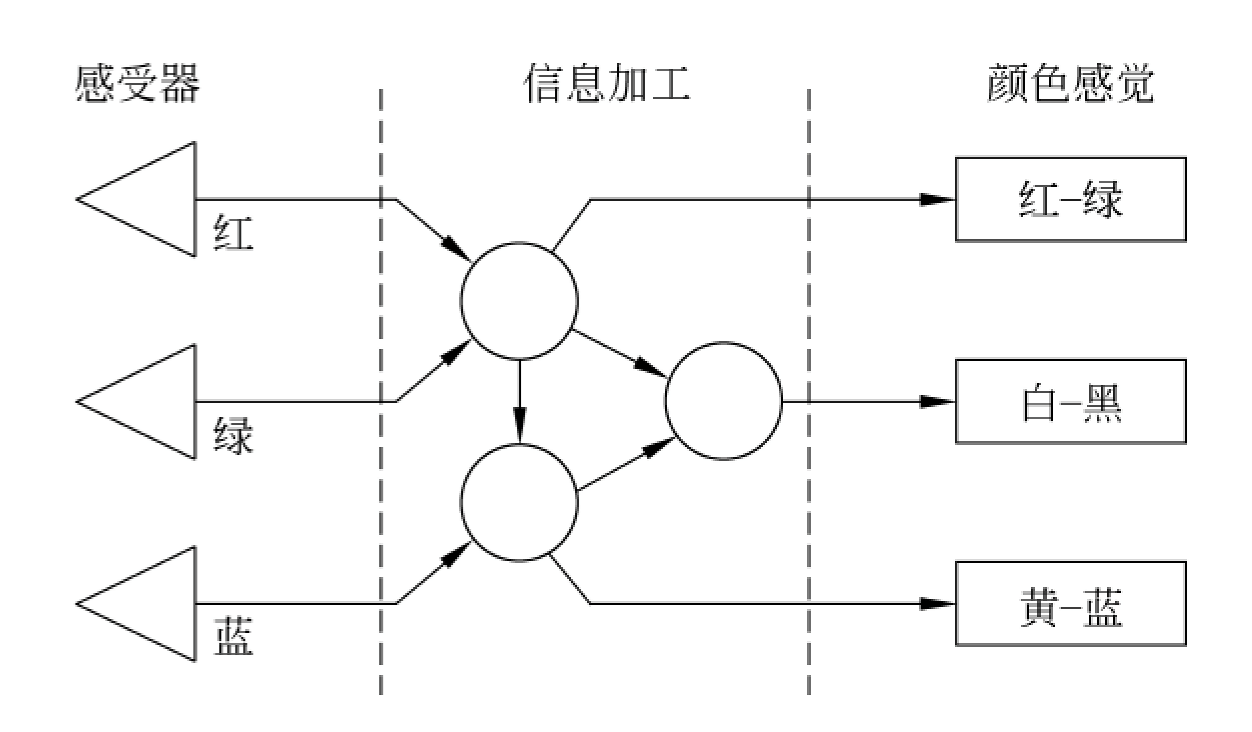



图 5.3　颜色视觉机制示意图

大小恒常性(Size Constancy)即大小知觉恒常性。人对物体的知觉大小不完全随视象大小而变化,它趋向于保持物体的实际大小。大小知觉恒常性主要是过去经验的作用,例如,同一个人站在离我们 3m、5m、15m、30m 的不同距离处,他在我们视网膜上折视象随距离的不同而改变着(服从视角定律)。但是,我们看到这个人的大小却是不变的,仍然按他的实际大小来感知。例如,在图 5.4 中,我们看到了庞佐错觉(Ponzo Illusion),图中央看起来大小不一的两个线条实际上是一样长的。庞佐错觉是因为两条趋近的线条造成了深度线索而产生的,不同深度的大小相同的图像通常显得大小不同。另外一种错觉是图 5.5 所示的缪勒-莱尔错觉,其中,长短相同的线条看起来长短不一。最后,请读者比较图 5.6 中央的圆圈。两个中央圆圈的大小实际上是一样的,但周围圆圈影响了对中央圆圈的知觉。

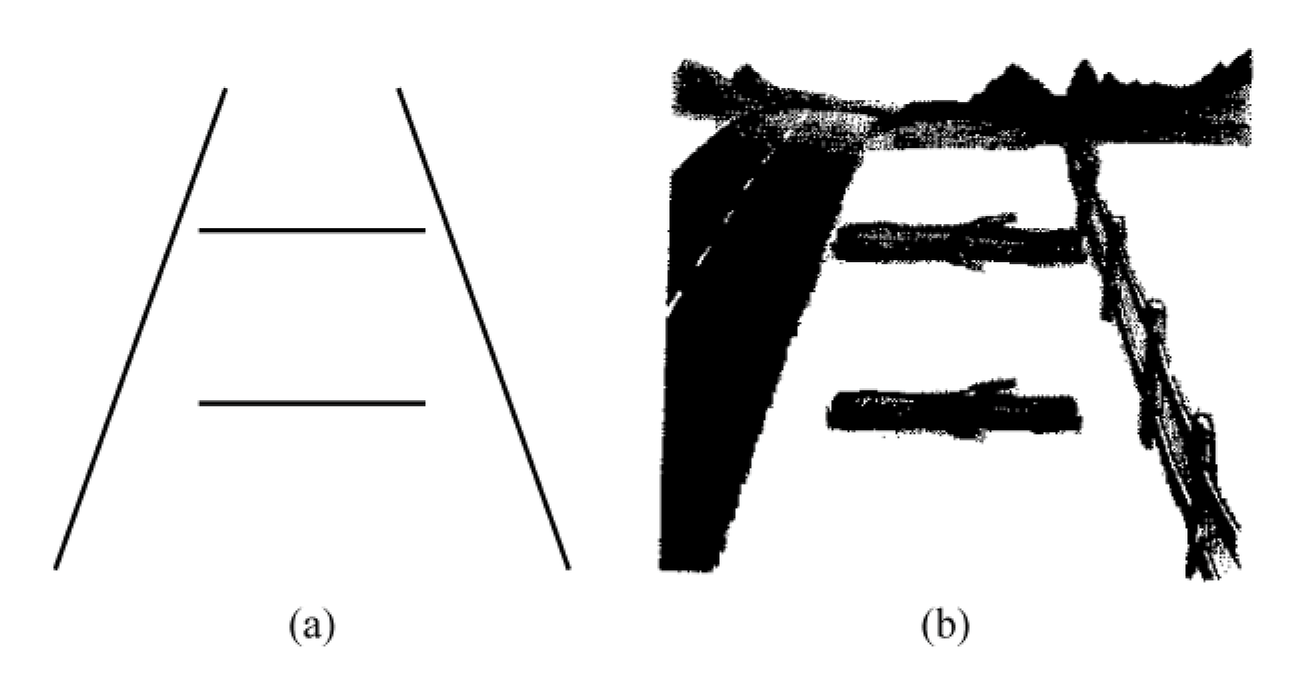



图 5.4　庞佐错觉

(知觉该图时,我们会认为图(a)中上面的线比下面的线长,图(b)中上面的木头比下面的木头长,尽管两条线和两根木头长短一样。这是因为在现实的三维世界中,上面的线和木头会更长。)

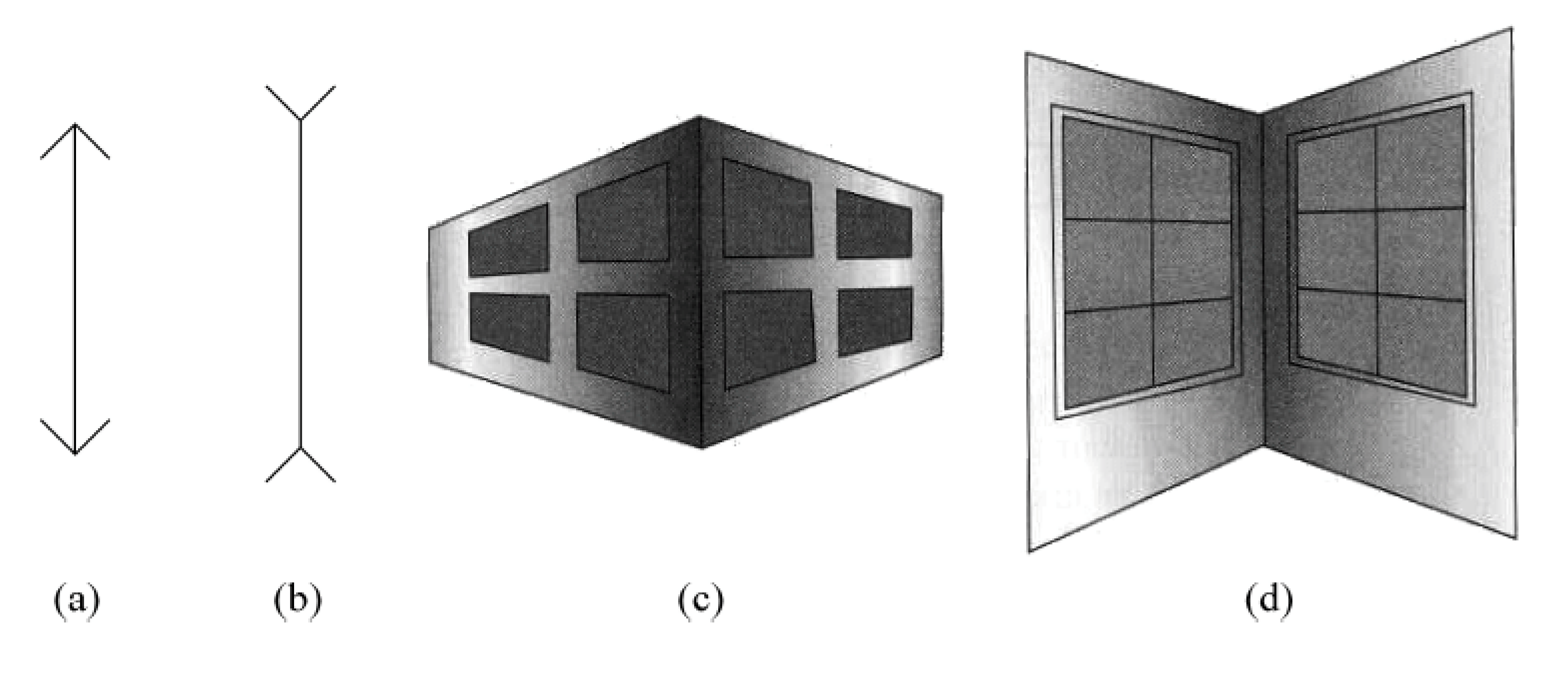



图 5.5　缪勒-莱尔错觉

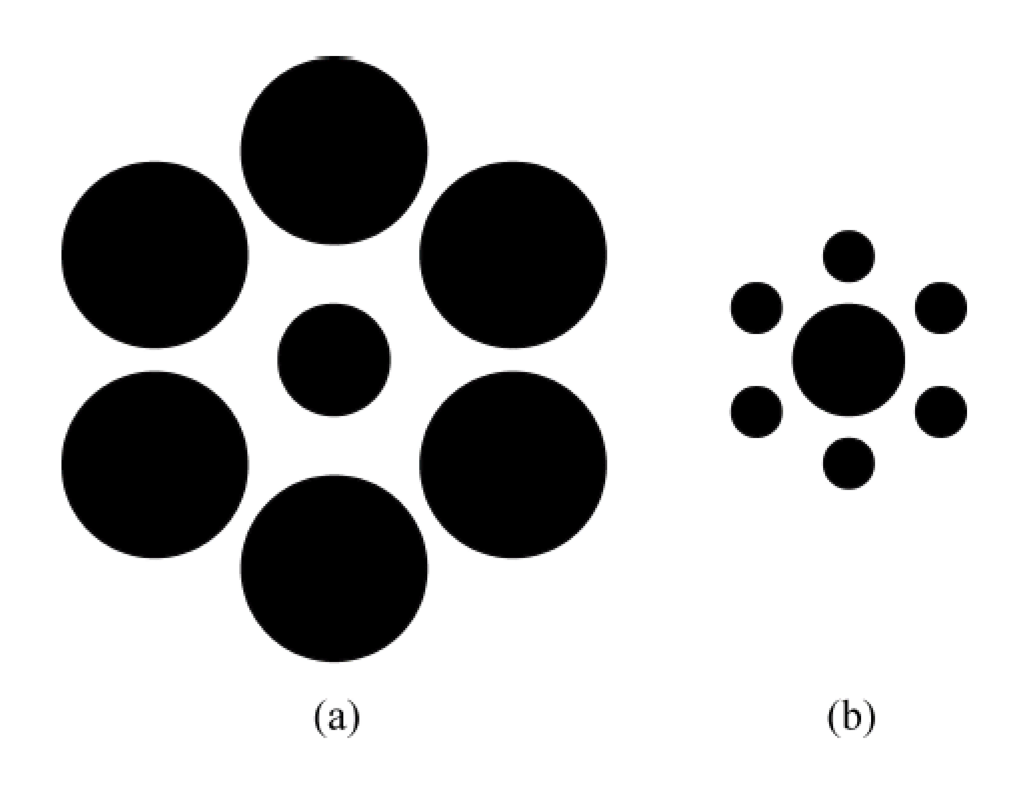



图 5.6 相对大小错觉

影响大小知觉常性的因素有：①刺激条件。条件越复杂，则越表现出常性，当刺激条件减少，则常性现象减少；②距离因素。距离很远时，常性消失；③水平观察时，常性表现大。垂直观察时，常性表现小。此外，在用人工瞳孔时，大小恒常性则消失。

形状恒常性(Form Constancy)即形状知觉恒常性。人从不同角度观察物体，或者物体位置发生变化时，物体在视网膜上的投射位置也发生了变化，但人仍然能够按照物体原来的形状来知觉(见图 5.7)。例如，房间门被打开时，它在视网膜上的视像形状与实际形状不完全一样，但看到门的形状仍是不变的。形状恒常性表明，物体的形状知觉具有相对稳定的特性。人的过去经验在形状恒常性中起重要作用。

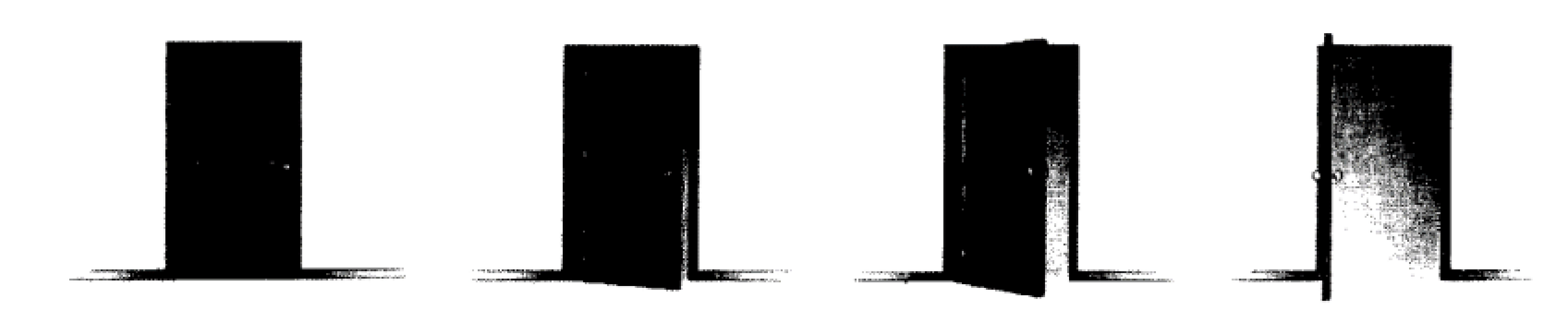


图 5.7 形状知觉

颜色恒常性(Color Constancy)即颜色知觉恒常性。在不同的照明条件下，人们一般可正确地反映事物本身固有的颜色，而不受照明条件的影响。例如，不论在黄光还是在蓝光的照射下，人们总是把红旗知觉为红色的，而不是黄色的或是蓝色的。黑林认为颜色知觉的恒常倾向是由于记忆色的影响。颜色恒常性可保证人对外界物体的稳定的辨认，具有明显的适应意义。

距离恒常性(Distance Constancy)又称距离的不变性，是指物体与知觉者的距离发生变化时，物体在网膜上造像的大小也发生相应的变化，但人知觉到的距离有保持原来距离的趋势的特性。

明度恒常性(Brightness Constancy)在不同照明条件下，人知觉到的明度不因物体的实际亮度的改变而变化，仍倾向于把物体的表面亮度知觉为不变。在强烈的阳光下煤块反射的光量远大于黄昏时白粉笔反射的光量，但即使在这种情况下，人们还是把煤块知觉为黑色的，把粉笔知觉为白色的。这就是明度恒常性现象。人们对物体亮度的知觉取决于它反射到眼中的光量，反射的光量愈大，就愈明亮。但在不同的照明条件下，物体的反射率(反射入射光量的百分比)是定常的。明度知觉恒常性是因人们考虑到整个环境的照明情况与视野

内各个物体反射率的差异,如果周围环境的亮度结构遭受不正常的变化,明度恒常性就会破坏。通常采用匹配法来研究明度常性,用邵勒斯比率来计算明度常性系数。

## 5.2 视觉理论

在视觉链中,大脑是否仅仅是视觉的被动场所,许多视觉研究者已经提出了不同的理论,建立了4种视觉理论。格式塔探讨偏重强调视觉组织的先天论的因素。构造论者的探讨对于学习和记忆的因素赋予较大的影响。动作探讨集中于视觉者在他的环境中做动作探测所产生的反馈作用。吉布森(Gibson)的生态学的探讨着重在刺激模式中所固有的全部环境的信息。

### 5.2.1 建构理论

过去的知识经验主要是以假设、期望或因式的形式在知觉中起作用。人在视觉时,接收感觉输入,在已有经验的基础上,形成关于当前的刺激是什么,或者激活一定的知识单元而形成对某种客体的期望。知觉是在这些假设、期望等的引导和规划下进行的。黑尔姆霍兹认为来自感觉的不充分信息可因无意识推论而得到增强,无意识推论可对感觉信息追加一层意义。在此基础上,布鲁纳等发展建构理论,认为所有感知都受到人们的经验和期望的影响[114]。建构理论的基本假设如下。

(1) 知觉是一个活动的、建构的过程,它在某种程度上要多于感觉的直接登记,其他事件会切入到刺激和经验之中来[277]。

(2) 知觉并不是由刺激输入直接引起的,而是所呈现刺激与内部假设、期望、知识以及动机和情绪因素交互作用的终极产品。

(3) 知觉有时可受到不正确的假设和期望的影响。因而,知觉也会发生错误。

建构理论关于知觉的看法是把记忆的作用赋予极大的重要性。他们认为先前经验的记忆痕迹,加到此时此地被刺激诱导出来的感觉中去,因此就构造出一个知觉象。而且,建构论者主张有组织的知觉基础是从一个人的记忆中选择、分析并添加刺激信息的过程,而不是格式塔论者所主张的大脑组织的天生定律所引起的自然操作作用。

已有的知识经验对知觉的影响是多方面的,最引人注目的是体现为上下文的作用。当前一些心理学家认为,总的看来,过去的知识经验主要是以假设、期望或因式的形式在知觉中起作用的。人在知觉时,接收感觉输入,在已有经验的基础上,形成关于当前的刺激是什么或者激活一定的知识单元而形成对某种客体的期望。知觉是在这些假设、期望等的引导和规划下进行的。依照格雷哥里(Gregory)的看法,知觉是一种包含假设考验的建构过程。人通过接收信息、形成和考验假设,再接收或搜寻信息,再考验假设,直至验证某个假设,从而对感觉刺激作出正确的解释,这被称作知觉的假设考验说。照这个学说看来,感觉刺激的物理特征、刺激的上下文和有关的概念都可激活长时记忆中的有关知识而形成各种假设。知觉因而是以假设为纽带的现实刺激信息和记忆信息相结合的再造。在通常情况下,人们在知觉时意识不到假设的参与,但在某些特殊条件下,如在弱的照明下看东西,有时是可以体验到这种假设考验的。假设考验说赋予知觉过程以主动性和智慧性的色彩,是目前在认

知心理学中有相当影响的知觉理论。

知觉的假设考验说是一种建立在过去经验作用基础上的知觉理论。支持这个理论的还有其他的重要论据。例如,外部刺激与知觉经验并没有一对一的关系,同一刺激可引起不同的知觉,不同的刺激却又可以引起相同的知觉。因此,感觉刺激的维量是模糊的,具有双关性质。感觉输入也是模糊的和片断的,不能对外部刺激提供真实而完整的描述。近端刺激如视网膜像只提供关于外部刺激的线索,需要在过去经验的基础上,应用这些线索作出推断,对近端刺激进行评价或解释,才能实现对外部刺激的知觉。这些看法显然是有利于假设考验说,也重视过去的知识经验在知觉中的作用。他们认为,知觉是定向、抽取特征,与记忆中的知识相对照,然后再定向、再抽取特征并再对照,如此循环,直到确定刺激的意义,这与假设考验说有许多相似之处。

### 5.2.2　直接知觉

美国心理学家吉布森因其对知觉的研究而闻名于学术界。1950 年他提出生态知觉理论[264],认为知觉是直接的,没有任何推理步骤、中介变量或联想。生态学理论(刺激物说)与建构理论(假设考验说)相反,主张知觉只具有直接性质,否认已有知识经验的作用。吉布森认为,自然界的刺激是完整的,可以提供非常丰富的信息,人完全可以利用这些信息,直接产生与作用于感官的刺激相对应的知觉经验,根本不需要在过去经验基础上形成假设并进行考验。根据他的生态知觉理论,知觉是和外部世界保持接触的过程,是刺激的直接作用。他把这种直接的刺激作用解释为感官对之做出反应的物理能量的类型和变量。知觉是环境直接作用的产物这一观点,是和传统的知觉理论相背离的。吉布森把表示一个人周围刺激的“刺激生态学”概念用公式来表示,它们包括倾斜的和反射的表面的关系,以及人们在走路、坐着和躺下时都感受到的引力,他坚信知觉不变,因此当环境提供给活跃的有机体连续而稳定的信息流时,有机体能够对此作出反应。这些观点反映在 1950 年出版的《视觉世界的知觉》和 1961 年出版的《作为知觉系统的感觉》[261]等著作中。

吉布森的知觉理论之所以冠之以“生态知觉理论”,原因在于它强调与生物适应最有关系的环境事实。对吉布森而言,感觉是因演进而对环境的适应,而且环境中有些重要现象,如重力、昼夜循环和天地对比等,在进化史上都是不变的。不变的环境带来稳定性,并且提供了个体生活的参照框架。因此,种系演化的成功依靠正确地反映环境的感觉系统。从生态学的观点来看,知觉是环境向知觉者显露的过程,神经系统并非建构知觉,而是萃取它们。在吉布森看来,感觉刺激所具有的丰富信息在实验室条件下被排除了许多,例如用速示器来迅速呈现刺激就是这样。但人在日常生活中,通常是有足够时间去观察的,而且人可以走动,从而改变观察的角度。随着观察点的改变,光流的一些特征发生变化,另一些特征则保持不变。像前面指出的那样,在空间的任一点上,都有其特定的光线分布,观察点的改变必然导致光线分布的变化,但光线分布总是含有一定结构。吉布森认为知觉系统从流动的系列中抽取不变性。他的理论现在称作知觉的生态学理论,并形成了一个学派,主要假设如下。

(1) 刺激眼睛的光线模式是一个光学分布(Optic Array);这种结构性的光线包含来自环境中的所有投射到眼睛的视觉信息。

(2) 这种光学分布提供关于空间中目标分布特征的明确的或恒定的信息。这种信息存

在多种形式,包括结构极差、光流模式和功能承受性。

(3) 知觉是在很少或没有信息加工参与的情况下,通过共振直接从光学分布中提取各种丰富信息。

吉布森把具有结构的表面的知觉叫做正常的或生态学的知觉。他认为,与他自己的看法相比,格式塔理论主要以特殊情况下的知觉分析为根据,在这种情况下,结构化减少了或者是毫不相干的,就像这张纸的结构与印在上面的内容毫不相干一样。

在构造论理论中,知觉常常是利用来自记忆的信息。而吉布森认为;具有结构表示的高度结构起来的世界提供了足够丰富而精确的信息,观察者可以从中选择,而无须再从过去储存起来的信息中选择。吉布森相信,通常是知觉指导我们的运动。这点与知觉的动作理论也不同。动作理论假设我们的肌肉所得到的信号变成了对知觉的影响,发展中的运动技巧指导发展中的知觉技巧。生态学理论坚信人们都是用相似的方法去看待世界,高度重视在自然环境中可得到的信息的全面复合的重要性。

吉布森的生态知觉理论具有一定的科学依据。他假设知觉反应是天生的观点与新生动物的深度知觉是一致的,同时也符合神经心理学中视觉皮层单一细胞对特定视觉刺激有所反应的研究结论。但是,他的理论过分强调个体知觉反应的生物性,忽视了个体经验、知识和人格特点等因素在知觉反应中的作用,因而也受到了一些研究者的批评。

建构理论与吉布森范式的区别之一是前者重视自上而下加工在知觉中的作用,而后者则强调自下而上加工的重要性。事实上,自上而下加工和自下而上加工对知觉的相对重要性取决于不同因素的影响。当观察条件良好时,视知觉主要由自下而上加工决定,但是当快速呈现刺激或刺激清晰度不够导致观察条件不理想时,视知觉主要涉及自上而下加工过程。与以上分析一致的是,吉布森重点考察优化条件下的视知觉,而建构主义则常常选用一些不太理想的观察条件来进行知觉研究。

间接和直接理论存在很大的区别,因为相关的理论家所追求的目标很不相同。如果我们考虑针对识别的知觉和针对行动的知觉之间的区别的话,这一点就会明朗得多。来自认知神经科学和认知神经心理学的证据也支持二者之间存在区别这一观点。这方面的证据表明一条腹侧加工通路更多地参与针对识别的知觉,而一条背侧加工通路更多地参与针对行动的知觉绝大多数知觉理论家都集中在探讨针对识别的知觉上,而吉布森则强调针对行动的知觉。

### 5.2.3 格式塔理论

格式塔(Gestalt)心理学诞生于 1912 年[807]。格式塔心理学家发现的感知组织现象是一种非常有力的关于像素整体性的附加约束,从而为视觉推理提供了基础。格式塔是德文 Gestalt 的译音。英文中常译成 Form(形式)或 Shape(形状)。格式塔心理学家所研究的出发点是“形”,它是指从由知觉活动组织成的经验中的整体。换言之,格式塔心理学家认为任何“形”都是知觉进行了积极组织或构造的结果或功能,而不是客体本身就有的。它强调经验和行为的整体性,反对当时流行的建构主义元素学说和行为主义“刺激-反应”公式,认为整体不等于部分之和,意识不等于感觉元素的集合,行为不等于反射弧的循环。尽管格式塔原理不只是一种知觉的学说,但它却导源于对知觉的研究,而且一些重要的格式塔原理,大

多是由知觉研究所提供的。

格式塔理论认为，知觉问题涉及比较和判断。当我们说这种灰色比那种灰色淡些，这根线条比那根线条长些，这个音比那个音响些时，我们所具有的经验究竟是什么呢？可以用一个实验予以阐释：在一块黑色平面上并排放着两个灰色小方块，要求判断两个灰色是否相同。回答有四种可能性：①在黑色平面上看见一大块颜色相同的灰色长方形，长方形中有一个分界线，将长方形分成两个方形；②看见一对明度梯度，从左至右上升，左边方形较暗，右边方形较亮；③看见一对相反方向的明度梯度，从左至右下降，左边方形较亮，右边方形较暗；④既未看见同色的长方形也未看见梯度，只有一些不确定的、模糊的东西。

从这些经验得出的判断是：①相同的判断；②左方形深灰色，右方形浅灰色；③左方形浅灰色，右方形深灰色；④不肯定或拿不准。

根据上述的描绘，在理论上可以推断出什么呢？格式塔心理学认为该描绘解释了两种比较的现象，“比较不是一种附加在特定感觉之上的新的意动……而是发现一个不可分的、联结着的整体。”以②和③的判断为例，梯度的意思并不只是指两个不同的层次，还指上升本身，即向上的趋势和方向，它不是一个分离的、飘忽的、过渡的感觉，而是整个不可分的经验的中心特征。

格式塔心理学相信大脑中组织的固有和天生的法则，在各种知觉因素之后存在着一个“简单性”原则，如果一个构造可以有一种以上的方式看到，那个“较简单的”方式会更通常一些。它虽然也重视经验对知觉的重要性，但首要着重点是放在神经系统的内在机制上的。

格式塔派学者们相信大脑中组织固有和天生的法则。他们辩论说，此法则就解释了这些重要现象：图形——背景的分化、对比、轮廓线、趋合、知觉组合的原则以及其他组织上的事实。格式塔派学者们认为，在他们所提出的各种知觉因素之后存在着一个“简单性”原则。他们断言，包含着较大的对称性、趋合、紧密交织在一起的单位以及相似的单位的任何模式，对于观察者来说，外表上显得“比较简单”。如果一个构造可以有一种以上的方式看到，例如，一个线条构成的图画可以看成是扁平的或者看成一个正方块，那个“较简单的”方式会更通常一些。格式塔派学者们并没有忽视潜在经验对于知觉的效应，但是他们的首要着重点是放在成为神经系统不可分的内在机制的作用上。因此，他们假设，似运动或Φ现象是大脑的天生的组织起来的倾向的结果。

单个图形背景的模式一般很少，典型的模式是几个图形有一个共同的背景。一些单个图形还倾向于被知觉集聚在一起的不同组合。格式塔心理学创始人之一的韦特海姆系统地阐述了如下“组合原则”。

(1) 邻近原则。彼此紧密邻近的刺激物比相隔较远的刺激物有较大的组合倾向。邻近可能是空间的，也可能是时间的。按不规则的时间间隔发生的一系列轻拍响声中，在时间上接近的响声倾向于组合在一起。由于邻近而组合成的刺激不必都是同一种感觉形式的；例如，夏天下雨时，雷电交加，我们就把它们感知为一个整体，即知觉为同一事件的组成部分。

(2) 相似原则。彼此相似的刺激物比不相似的刺激物有较大的组合倾向。相似意味着强度、颜色、大小、形状等这样一些物理属性上的类似。俗话说，“物以类聚，人以群分”，也就包含这种原则。

(3) 连续原则。人们知觉倾向于知觉连贯或连续流动的形式，即一些成分和其他成分连接在一起，以便有可能使一条直线、一条曲线或者一个动作沿着已经确立的方向继续下去。

(4) 闭合原则。人们知觉倾向于形成一个闭合或更加完整的图形。

(5) 对称原则。人们知觉倾向于把物体知觉为一个中心两边的对称图,导致对称或平衡的整体而不是非对称的整体。

(6) 共方向原则。也有称共同命运原则。如果一个对象中的一部分都向共同的方向去运动,那这些共同移动的部分就易被感知为一个整体。这个组合原则本质上是相似组合在运动物体上的应用,它是舞蹈设计中的一个重要手段。

在每一种刺激模式中,一些成分都有某种程度的接近、某种程度的类似以及某种程度适合"好图形"的东西。有时组合的一些倾向在同一方向上起作用,有时它们彼此冲突。例如,图 5.8 给出了格式塔知觉组织原则例图。

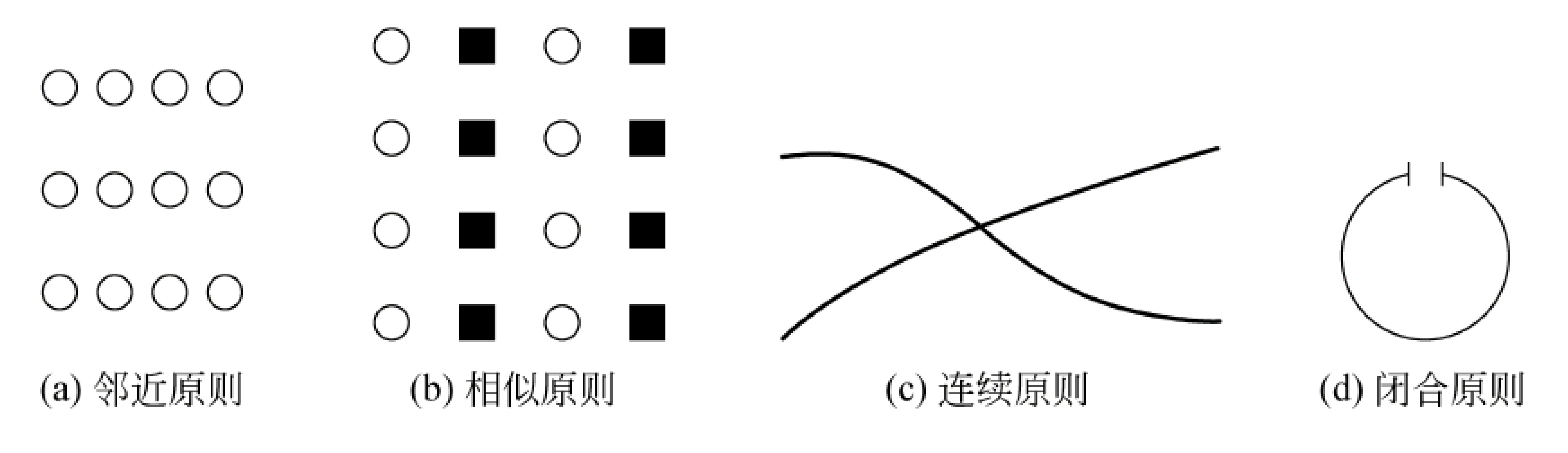



图 5.8　格式塔知觉组织原则例图

格式塔心理学家试图根据心脑同形观来解释知觉原则。按照这种心脑同形观,视觉组织经验与大脑中的某一过程严格对应。当我们观察环境时,格式塔心理学家假定大脑中存在一种电场,以帮助产生相对稳定的知觉组织经验。格式塔心理学家主要依赖内省报告或"注视一个图形并从自己的角度观看"的方法研究知觉。不幸的是,格式塔心理学家对大脑的工作机制知之甚少,而且他们的虚拟生物学解释也没有得到承认。

格式塔理论反映了人类视觉本质的某些方面,但它对感知组织的基本原理只是一种公理性的描述,而不是一种机理性的描述。因此自从 20 世纪 20 年代提出以来未能对视觉研究产生根本性的指导作用,但是研究者对感知组织原理的研究一直没有停止。特别是在 20 世纪 80 年代以后,威特肯(Witkin)和坦丁鲍姆(Tenenbaum)、劳卫(Lowe)、蓬特兰德(Pentland)等人在感知组织的原理,以及在视觉处理中的应用等方面取得了新的重要研究成果。

## 5.3　视觉有效编码

人们一致认为,初级视皮层的视觉处理过程受环境统计特性的影响,但是怎样在两者之间建立准确的数字连接一直是一个难题。在信息理论的影响和指导下,阿特尼夫(F Attneave)于 1954 年提出视觉感知的目标就是产生一个外部输入信号的有效表示[38]。在神经生物学领域,巴罗(H B Barlow)提出了"有效编码假设"[57],认为初级视皮层神经细胞的主要功能就是去除输入刺激的统计相关性。如果这种有效编码假设是正确的,对于神经细胞的响应特性,我们可以期望观察到什么样的行为?神经系统通过两种方式体现这种有效编码特性:单个神经细胞响应的分布曲线和多个神经细胞间的统计独立性。

考察单个神经细胞对自然环境响应的分布状态时,为了检测单个神经细胞是否最大化

的表达了输入信号信息，需要对神经细胞响应值添加一些限制条件。例如，我们假定神经细胞响应值存在最大值 Rmax，很容易得到，当神经细胞的响应在[0,Rmax]满足平均分布时可使得信息最大化，也就是，神经细胞必须均等地使用在该区间中的每一个可能值。神经细胞响应的最优分布紧密依赖神经细胞响应约束条件。例如，选择约束条件为神经细胞响应的方差恒定，那么信息最大化的分布则对应高斯分布；如果约束条件是均值恒定，那么信息最大化的分布对应指数分布。

如果一组神经细胞联合编码一个外部刺激模式，除了单个神经细胞有效编码外部刺激外，多个神经细胞的联合响应也必须尽可能大的编码输入刺激，显然，如果某一个信息被多个神经细胞编码，那么这组神经细胞的联合响应效率就会下降。从数学意义上，如果每个神经细胞的响应满足统计独立条件，那么这组神经细胞的编码是最有效的。两个神经细胞响应满足独立性是指，不能通过一个神经细胞的响应来预测另一个神经细胞的响应，多神经细胞的有效编码与单神经细胞的有效编码相比，这里的独立性目标不需要任何附加约束条件。

从方法论上，有效编码研究主要可以分为两条路线：直接的方法是从生物机理上，在自然图像刺激条件下检测神经细胞的响应特性，这被称为机理测试；另一个替代的方法是利用自然图像的统计特性，建立模型模拟早期视觉处理系统的处理机制，这被称为模型仿真。

实验可以检测在自然图像或者图像序列刺激下，视神经细胞的响应特性，这些研究用不同的方法测试视神经细胞编码的有效性，绝大部分的试验结果验证了有效编码假设。巴德列(R Baddeley)等在猫的 V1 区细胞和猴子的 IT 区细胞进行试验，发现在自然图像刺激下这些神经细胞的发放率(Firing Rate)的分布服从指数分布，这与单个神经细胞有效编码的准则一致，即在固定平均发放率的条件下，指数分布传送的信息最多[54]。尼伦伯格(S Nirenberg)等 2001 年在 Nature 上发表的研究结果表明，在冗余性测度和自然刺激条件下，一组视网膜神经节对外界刺激独立编码[542]。

神经生理学的研究表明，每个神经细胞都携带与统计无关的信息成分。在视网膜和侧膝状体神经元之间，在神经束的有限带宽条件下，视觉系统通过最优化信息传输来编码外部信息；然后，V1 区的神经细胞把传入的最优化刺激转换为稀疏编码，而且这些神经细胞表示着自然场景的独立信息成分。这种稀疏编码将促进对视觉系统高级视觉区域的理解，也能增加模式识别的有效性。

在模型仿真方面首先考察外部环境刺激的统计特性，根据特定的统计优化准则建立响应模型，这里叫做有效编码模型，然后与初级视皮层中神经细胞的响应特性(神经生理学)进行比较。下面重点研究怎么建立这种模型，形成外部刺激统计特性与视觉系统处理机制的联系。

在这里有效编码模型只针对多个神经细胞，而单个神经细胞的编码不在我们研究的范围之内。有效编码模型可以描述为：分解输入信号，形成一组独立的响应。有效编码的基本模型可以用如图 5.9 所示。图 5.9 中的每一个小的图像块对应于一个神经细胞的感受野，把它按列排列形成一个 $\boldsymbol{N}$ 维列向量 $\boldsymbol{X}$。假设 $\boldsymbol{X}$ 可以用 $\boldsymbol{N}$ 个基函数线性叠加而得到，这 $\boldsymbol{N}$ 个基函数构成了矩阵 $\boldsymbol{A}$ 的列，同时，线性叠加时每个基函数对应的权值用向量 $\boldsymbol{S}$ 表示，$\boldsymbol{S}$ 的每一个分量我们通常叫做“因子”。图像的线性叠加可以表示为

$$\boldsymbol{X} = \boldsymbol{AS} \tag{5.1}$$

有效编码模型就是对图像 $\boldsymbol{X}$ 进行变换(线性或者非线性)，找到隐藏的“因子”$\boldsymbol{S}$，并且使

得 $\boldsymbol{S}$ 满足有效性编码准则,例如统计独立性条件。简单的这个模型可表示为

$$\boldsymbol{S} = F(\boldsymbol{WX}) \tag{5.2}$$

这个简化的模型可以分为两个步骤,第一步是线性滤波,$\boldsymbol{W}$ 的列向量对应一个个线性滤波器;然后在线性滤波的结果上进行非线性的变换,这个步骤是可选的。

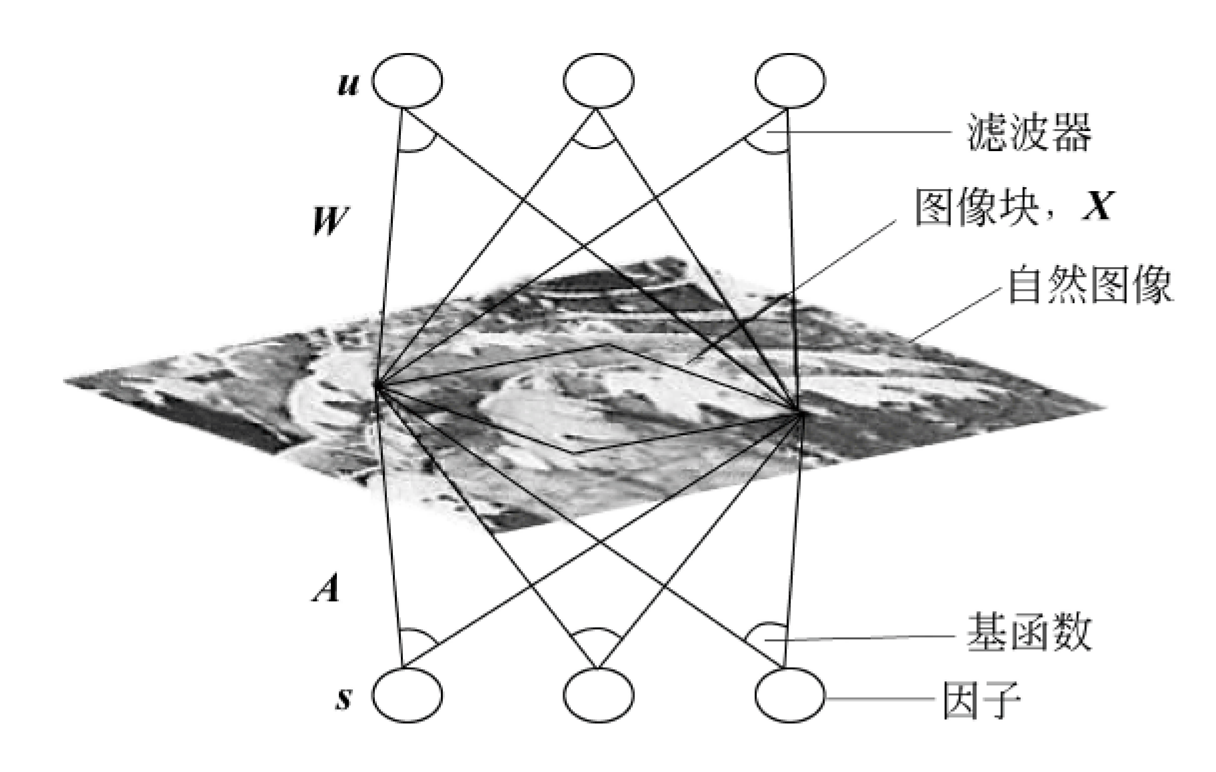



图 5.9 图像线性叠加和有效编码模型

随着输入信号维数的增加,描述输入信号联合分布的复杂度会指数增长。因此,通常针对特定的统计特性,或者指定特定的分解算法来简化问题。常用的有效编码算法可以分为三种。第一种是基于二阶统计量的线性去相关方法,对应图 5.10 中 $R_2$ 的对应的虚线部分。第二种是基于高阶统计量的独立成分分析方法,是图 5.10 中 $R_1$ 对应的实线部分;第三种是基于高阶统计量的非线性分解方法,对应图 5.10 中的 $R_3$。

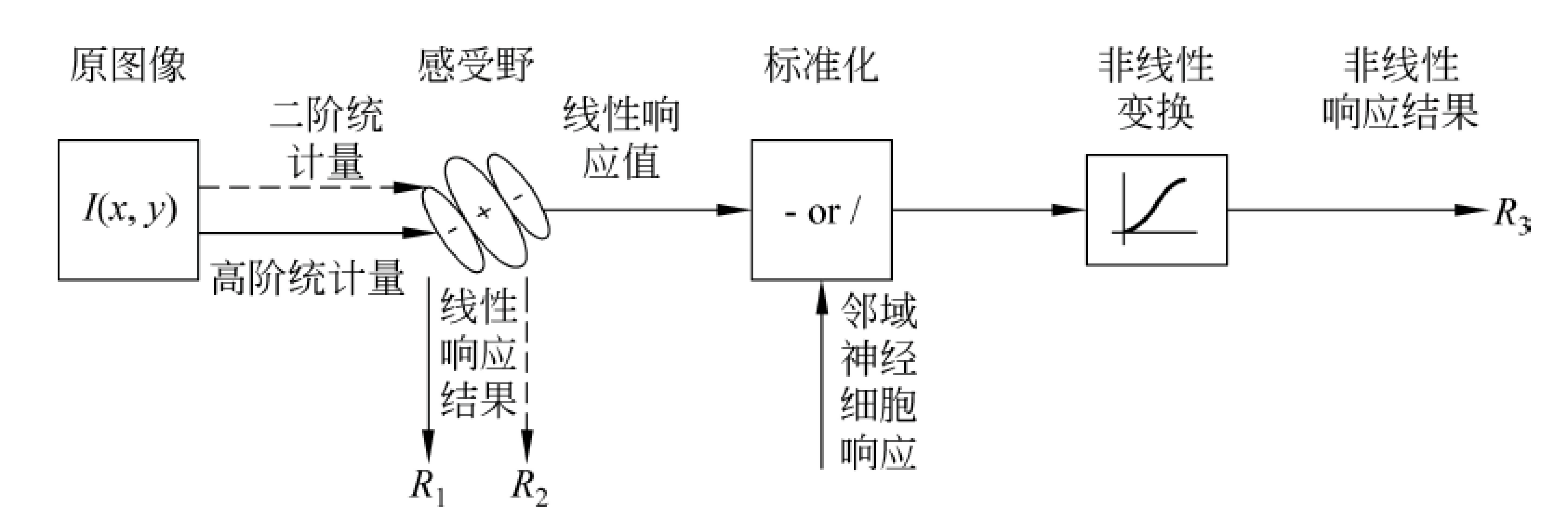



图 5.10 有效编码建模框图

### 1. 基于二阶统计的线性滤波器模型

最简单的约束就是考虑输入信号的二阶统计量(协方差、相关系数等),用线性分解进行建模,如图 5.10 中 $R_2$ 所对应的路线。主成分分析(PCA)可以解决这个问题。对应式(5.2)中 $F$ 为恒等变换,且使得 $\boldsymbol{S}$ 满足:

$$\mathrm{Cov}(\boldsymbol{S}) = \text{Diagonal Matrix} \tag{5.3}$$

主成分对应着一些这样的轴,沿着这些轴的方向可以消除各个数据维的相关性。我们可以严格证明,这些轴是肯定存在的,虽然他们的方向不一定唯一。尤其是,如果输入信号满足高斯分布,那么经过主成分分析产生的各个分量,或者各个轴的变量是相互独立的。

尽管 PCA 对于高斯分布的信号能够产生统计独立的分量或者轴,但是对于非高斯分布则是常常会失败。自然图像服从非高斯分布,因此用 PCA 分解后得到的“因子”满足统计无

关性,但是不能达到统计独立的要求。而 PCA 方法所对应的基函数反应的是图像的全局空间频率,如图 5.11 所示。

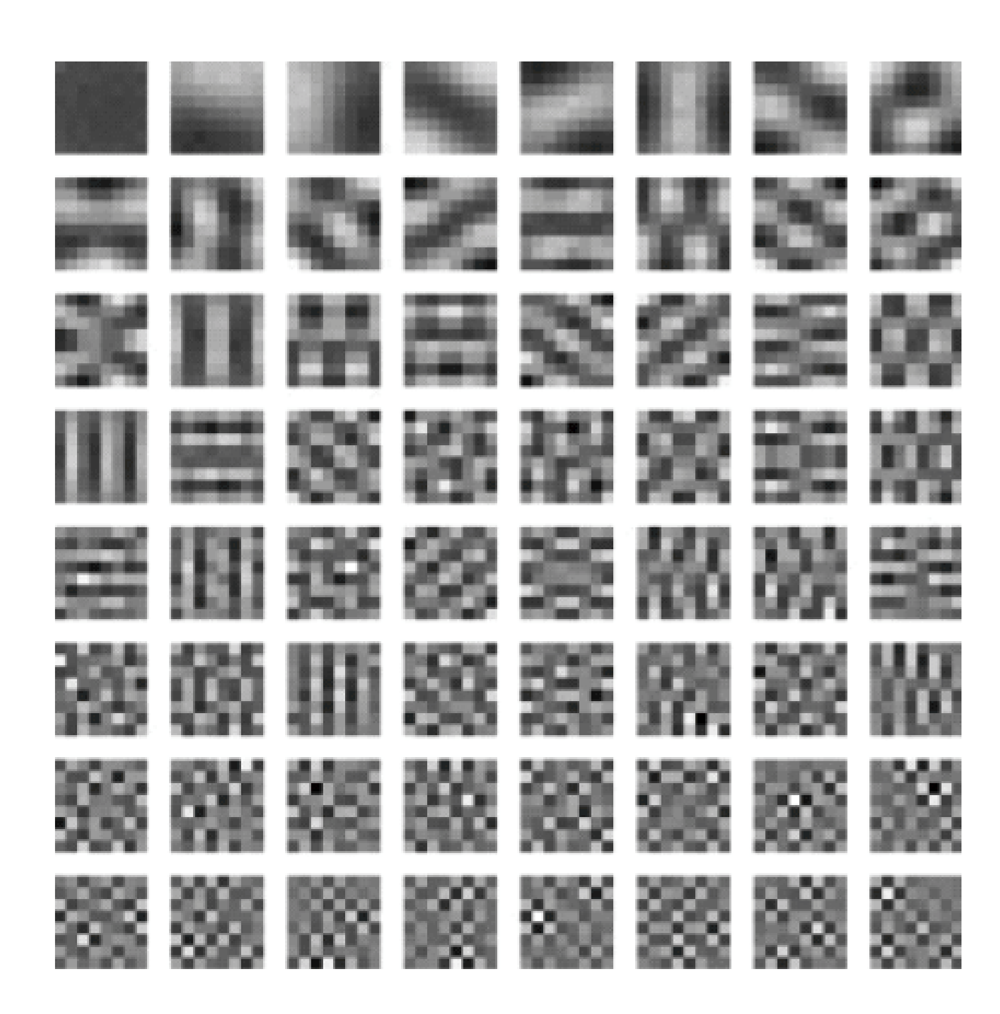


图 5.11　在 8×8 的图像块集合上 PCA 得到的基函数

## 2. 基于高阶统计的线性滤波器模型

PCA 得到主成分对应着图像的空间频率,这种空间频率或者说二阶统计量不足以反映图像的空间局部信息,比如边界轮廓。因此我们必须引入另外的约束准则。

费尔德(D J Field)提出了稀疏编码或者信息熵最小化编码的方法,也就是对应于一个输入图像刺激,激活的神经细胞的个数尽可能少[223]。通过对 Gabor 滤波器的参数进行最优化,他们发现这些优化参数产生的滤波器与视皮层中简单细胞的响应特性非常相似,它们具有位置选择性、方向选择性和频率选择性等特点。

奥尔肖森(B A Olshausen)等用线性叠加的方法进一步研究了简单细胞感受野性质与稀疏编码的关系。他们利用如式(5.1)所示的基函数线性叠加表示输入刺激图像,在最小均方差意义下使得线性叠加的结果尽可能的与原图像相似,也就是使重构的误差尽可能小,同时,使得表示特征尽可能稀疏化,或者说对应基函数权值为 0 的个数尽可能多。这个优化准则可表示为

$$E(a,\Phi)=\sum_{x,y}\left[I(x,y)-\sum a_i\Phi_i(x,y)\right]+\lambda_s\sum_i \boldsymbol{S}\left(\frac{a_i}{\sigma_i}\right) \tag{5.4}$$

式(5.4)对应式(5.1)中 $\boldsymbol{A}$ 的列向量。式中第一项表示基函数集重构原始图像完备的程度,用误差的平方和表示;第二项表征编码的稀疏特性,使得尽可能小的基函数表示原始图像。从自然图像中任意选择图像块构成一个训练集合,随机初始化基函数集合,经过几百次训练后得到的基函数具有与视皮层简单细胞类似的性质。

## 3. 非线性模型

斯瓦兹(O Schwartz)等通过在一个非线性变换,消除在线性编码结果的“因子”间的非线性相关性。首先调整每个基函数的响应值,典型的如平方运算,然后除以调整后的邻域神经细胞响应的加权和,得到新的响应值当作最后的神经细胞响应。这种调整对应到式(5.2)

中的函数 $F$。这个非线性过程可以看作是一个两阶段组成的层次模型:

$$\{x_i\} \xrightarrow{T} \{s_i\} \xrightarrow{R} \{r_i\} \tag{5.5}$$

原始图像像素点 $x_i$ 首先经过一个线性变换 $T$,得到线性滤波系数 $c_i$,然后,再对线性滤波系数进行非线性的变换 $R$。这里的线性滤波器就是具有位置、方向和频率选择性的滤波器。非线性变换 $R$ 是一种信息增益控制机制,通过在邻域线性滤波系数的能量加权和进行规范化每一个线性滤波系数:

$$ri = \frac{\text{sgn}(s_i) * \mid s_i \mid^{\gamma}}{\beta_i + \sum_j h_{ij} \mid s_j \mid^{\gamma}} \tag{5.6}$$

结果表明,经过这种非线性变换的编码系数能显著提高相邻滤波器系数之间的统计独立性,而这种相邻滤波器响应稀疏之间的相关性是用线性变换无法消除的。

这种非线性处理过程在大脑皮层中随处可见,而且类似的“分散规范化”模型在解释神经细胞非线性行为时得到广泛的使用。这种非线性关系能很好地适应自然图像的非高斯统计特性,而且,可以通过优化规范化过程中邻域的权值,使得神经细胞响应的统计独立性更大化。这种模型的编码结果与很多神经生理学数据具有惊人的一致性。

## 5.4 马尔的视觉计算理论

视觉信息处理是一个根据图像发现周围景物中有什么物体和物体在什么地方的过程,也就是从图像得到对观察者有用的符号描述的过程。从输入图像到得出景物描述之间存在着巨大的间隙,需要经过一系列的信息处理和理解过程。对这个过程本质的认识是揭开视觉之谜的关键,但目前我们对这些还远未了解清楚。

通过视觉识别物体就是把图像的元素与已知的景物中物体的描述或模型之间建立对应关系。图像中的元素是点状的像素,像素的值就是这个像素的灰度值,这是点状的数据。而与此相对,物体是通过它的形状、大小、几何结构、颜色等特征来描述的。这些特征代表物体的整体性质。要在输入的点状数据与物体的整体性质之间建立对应关系就必须要经过一个把点状数据组织(Grouping)起来的过程。这样的组织过程不只是在视觉中有,而且在听觉及其他感觉中也存在。

与如何形成整体性质相联系的问题是恒常性问题。大家都知道,图像中各点的灰度是景物中多种因素综合作用的结果。这些因素包括光照条件、物体表面的反射特性、观察者相对于物体的距离和方位、物体表面形状等。这些因素的任何变化都会改变图像的灰度,也就会改变我们看到的图像。但是我们通过视觉所感觉到的物体的形状、大小和颜色都是与观察者的状况以及照明条件无关的。具体而言,当照明条件和观察者相对物体的距离方位发生变化时,虽然在视网膜上产生的图像要随之而变化,但人看到的总是某种形状和大小的物体。但人在大脑中感觉到的是物体可变的外表后面的恒定特征。因此,大脑不但把点状的传感信息组织成整体,而且经过一个因素分解过程(Factoring)把这些影响传感器信息的条件,即照明条件、观察者的距离和方位等因素分离出去,得到纯粹的关于物体的信息。这些信息是不随上述条件而改变的,因此被称为具有恒常性(Constancy)。总之,大脑不是直接根据外部世界在视网膜上的投影成像,而是根据经过聚集过程和因素分解过程处理后的信

息来识别物体。

上述两种处理过程有一个非常重要的问题，是把点状的图像信息变换成整体描述的组织过程与对各种影响成像结果的因素进行分解过程之间的关系。在没有完成因素分解过程以前，我们能着手进行聚集过程吗？以美国麻省理工学院人工智能实验室马尔（D Marr）为首的一些科学家认为在得到关于物体的纯粹信息，例如深度、表面、方向、反射率等以前，做任何组织的处理都是无用的[462]。他们把这样的纯粹信息称为本征图像，因此他们采用基于重构（Reconstruction）的视觉信息处理方法，也就是通过重构这些本征图像来识别物体。而另一派科学家则认为某些预先进行的组织过程不仅可以为因素分解过程提供必要的基础，而且还可形成某种反应物体空间结构的图像关系，根据这些图像关系可以产生对图像内容的假设。因此，他们采用了基于推理和识别的视觉信息处理方法。前一种观点是以马尔关于人类视觉的计算理论为代表；后一种观点是以格式塔学派，及其后续者，如 Lowe、Pentland 等关于感知组织（Perception Organization）的理论为代表。这两派理论各自反映了视觉过程中的基本矛盾，但都未能对视觉过程作出满意的解释。这两种理论的争论推动了对视觉的研究（见图 5.12）。

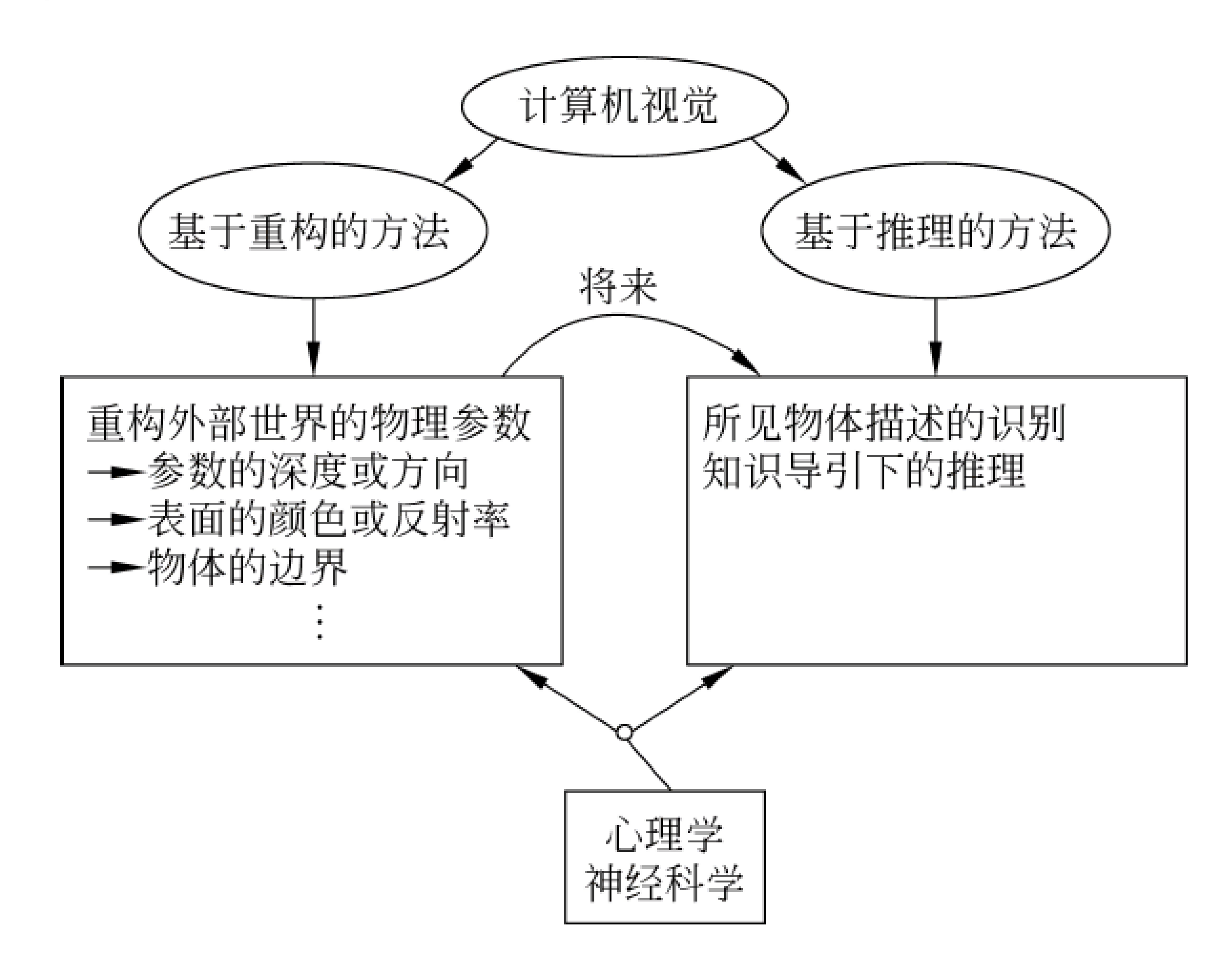



图 5.12 计算机视觉中的两种方法

马尔教授在 20 世纪 70 年代末，20 世纪 80 年代初创立了视觉的计算理论，使视觉的研究前进了一大步[462]。马尔的视觉计算理论立足于计算机科学，系统地概括了心理物理学、神经生理学、临床神经病理学等方面已取得的所有重要成果，是迄今为止最系统的视觉理论。马尔理论的出现对神经科学的发展和人工智能的研究产生了深远的影响。

马尔认为视觉是一个信息处理过程。这个过程根据外部世界的图像产生对观察者有用的描述。这些描述依次由许多不同但固定的、每个都记录了外界的某方面特征的表象（Representation）所构成或组合而成。一种新的表象之所以提高了一步是因为新的表象表达了某种信息，而这种信息将便于对信息作进一步解释。按这种逻辑来思考可得到这样的结论：在对数据作进一步解释以前我们需要关于被观察物体的某些信息，这就是所谓的本征图像。然而，数据进入我们的眼睛是要以光线为媒介的。灰度图像中至少包含关于照明情况、观察者相对于物体位置的信息。因此，按马尔的方法首先要解决的问题是如何把这些

因素分解开。他认为低层视觉(即视觉处理的第一阶段)的目的就是要分清哪些变化是由哪些因素引起的。大体上来说这个过程要经过两个步骤来完成:第一步是获得表示图像中变化和结构的表象。这包括检测灰度的变化、表示和分析局部的几何结构、以及检测照明的效应等处理。第一步得到的结果被称为初始简图(Primal Sketch)的表象;第二步对初始简图进行一系列运算得到能反映可见表面几何特征的表象,这种表象被称为二维半(2.5 D)简图或本征图像。这些运算中包括由立体视觉运算提取深度信息,根据灰度影调、纹理等信息恢复表面方向,由运动视觉运算获取表面形状和空间关系信息等。这些运算的结果都集成到本征图像这个中间表象层次。因为这个中间表象已经从原始的图像中去除了许多的多义性,是纯粹地表示了物体表面的特征,其中包括光照、反射率、方向、距离等。根据本征图像表示的这些信息可以可靠地把图像分成有明确含义的区域(这称为分割),从而可得到比线条、区域、形状等更为高层的描述。这个层次的处理称为中层视觉处理(Intermediate Processing)。马尔视觉理论中的下一个表象层次是三维模型,它适用于物体的识别。这个层次的处理涉及物体,并且要依靠和应用与领域有关的先验知识来构成对景物的描述,因此被称为高层视觉处理。

马尔的视觉计算理论虽然是首次提出的关于视觉的系统理论,并已对计算机视觉的研究起了巨大的推动作用,但还远未解决人类视觉的理论问题,在实践中也已遇到了很大的困难。对此现在已有不少学者提出改进意见。

马尔首先研究了解决视觉理解问题的策略。他认为视觉是一个信息处理问题。它需要从以下三个层次来理解和解决。

(1) 计算理论层次——研究对什么信息进行计算和为什么要进行这些计算。

(2) 表示和算法层次——实际计算由计算理论所规定的处理,输入输出如何表示好以及将输入变换到输出的算法。

(3) 硬件实现——实现由表示和算法级所考虑的表示,实现执行算法,研究完成某一特定算法的具体机构。

例如,傅里叶变换是属于第一层的理论,而计算傅里叶变换的算法,如快速傅里叶变换算法是属于第二个层次的。至于实现快速傅里叶算法的阵列处理机就属于硬件执行的层次。

可以认为视觉是一个过程,这个过程从外部世界的图像产生对观察者有用的描述。这些描述依次地由许多不同的、但是固定的、每个都记录了景物的某个方面的表示法所构成或组合而成。因此选择表示法对视觉的理解是至关重要的。根据马尔所提出的假设,视觉信息处理过程包括三个主要表示层次:初始简图、二维半简图和三维模型。根据某些心理学方面的证据,人类视觉系统的表示法如图5.13所示。

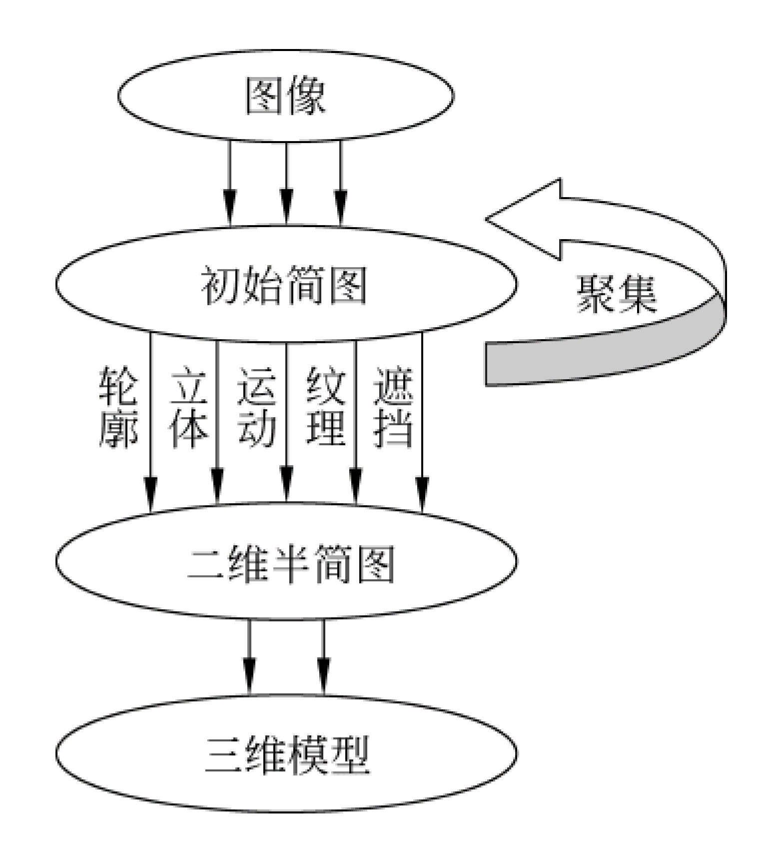


图5.13 视觉系统的表示层次

## 1. 初始简图

在灰度图像中,包含两种重要的信息:图像中存在的灰度变化和局部的几何特征。初

始简图是一种基元表示法，它可以完全而清楚地表示这些信息。初始简图所包含的大部分信息集中在与实际的边缘以及边缘的终止点有关的急剧的灰度变化上。每个由边缘引起的灰度变化，在初始简图上都有相应的描述。这样的描述包括：与边缘有关的灰度变化率，总的灰度变化、边缘的长度、曲率以及方向。粗略地说，初始简图是以勾划草图的形式来表示图像中的灰度变化。

### 2. 二维半简图

图像中的灰度受多种因素的影响，其中主要包括光照条件、物体几何形状、表面反射率以及观察者的视角等。因此，先要分清上述因素的影响，也就是对景物中物体表面作更充分的描述，才能着手建立物体的三维模型，这就需要在初始简图与三维模型之间建立一个中间表示层次，即二维半简图。物体表面的局部特性可以用所谓的内在特性来描述。典型的内在特性包括表面方向、观察者到表面的距离，反射和入射光照、表面的纹理和材料特性。内在图像由图像中各点的某项单独的内在特性值，以及关于这项内在特性在什么地方产生不连续的信息所组成(见表 5.1)。二维半简图可以看成是某些内在图像的混合物。简而言之，二维半简图完全而清楚地表示关于物体表面的信息。

在初始简图和二维半简图中，信息经常是以和观察者联系在一起的坐标为参考表示的，因此这种表示法被称为是以观察者为中心的表示法。

**表 5.1　二维半简图**

| 信　息　源 | 信 息 类 型 | 信　息　源 | 信 息 类 型 |
|---|---|---|---|
| 立体视觉 | 视差，因而可得到 $\delta\gamma$、$\Delta\gamma$ 和 $S$ | 其他遮挡线索 | $\Delta\gamma$ |
| 方向选择性 | $\Delta\gamma$ | 表面方向轮廓 | $\Delta s$ |
| 从运动恢复结构 | $\gamma$、$\delta\gamma$、$\Delta\gamma$ 和 $S$ | 表面纹理 | 可能有 $\gamma$ |
| 光源 | $\gamma$ 和 $S$ | 表面轮廓 | $\Delta\gamma$ 和 $S$ |
| 遮挡轮廓 | $\Delta\gamma$ | 影调 | $\delta s$ 和 $\Delta s$ |

注：$\gamma$ 相对深度(按垂直投影)，就是观察者到表面点的距离；$\delta\gamma$、$\gamma$ 的连续或小的变化；$\Delta\gamma$，$\gamma$ 的不连续点；$S$ 局部表面方向；$\delta s$，$S$ 的连续或小的变化；$\Delta S$，$S$ 的不连续点。

在初始简图和二维半简图中，信息经常是以和观察者联系在一起的坐标为参考表示的，因此这种表示法被称为是以观察者为中心的表示法。

### 3. 三维模型

在三维模型表象中，以一个形状的标准轴线为基础的分解最容易得到。在这些轴线中，每条轴线都和一个粗略的空间关系相联系；这种关系对包含在该空间关系范围内的主要的形状组元轴线提供了一种自然的组合方式。用这种方法定义的模块我们称之为三维模型。所以，每一个三维模型说明[462]如下几点。

(1) 一根模型轴，指的是能确定这一模型的空间关系的范围的单根轴线。它是表象的一个基元，能粗略地告诉我们被描述的整体形状的若干性质，例如，整体形状的大小信息和朝向信息。

(2) 在模型轴所确定的空间关系机含有主要组元轴的相对空间位型和大小尺寸可供选

择。组元轴的数目不宜太多,它们的大小也应当大致相同。

(3) 一旦和组元轴相联系的形状组元的三维模型被构造出来,那么就可以确定这些组元的名称(内部关系)。形状组元的模型轴对应于这个三维模型的组元轴。

在图5.14中[462],每一个方框都表示一个三维模型,模型轴画在方框的左侧,组元轴则画在右侧。人体三维模型的模型轴是一基元,它把整个人体形状的大体性质(大小和朝向)表达清楚。对应于躯干、头部、肢体的六根组元轴各自可以和一个三维模型联系起来,这种三维模型包含着进一步把这些组元轴分解成更小的组元构型的附加信息。尽管单个三维模型的结构很简单,但按照这种层次结构把几个模型组合起来,就能在任意精确的程度上构成两种能抓住这一形状的几何本质的描述。我们把这种三维模型的层次结构称为一个形状的三维模型描述。

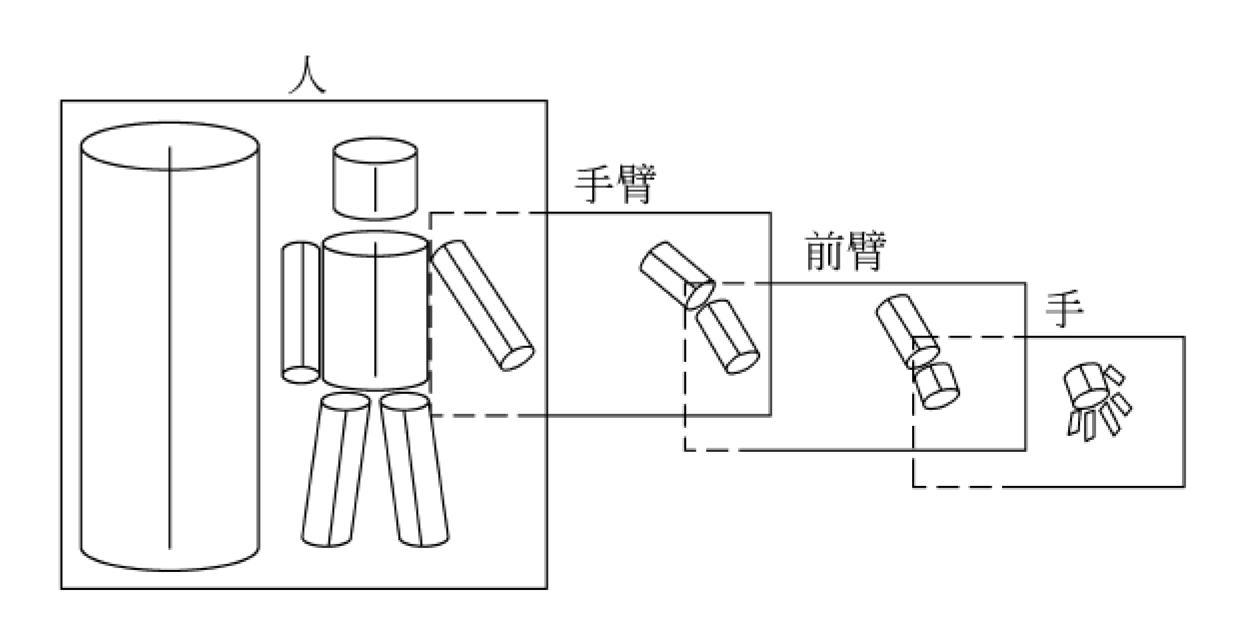



图5.14 人的三维模型

三维表示法完全而清楚地表示有关物体形状的信息。采用广义柱体的概念虽然很重要,却很简单。一个普通的圆柱可以看成是一个圆沿着通过它的中心线移动而形成的。更一般的情况,一个广义柱体是二维的截面沿着称为轴线移动而成。在移动过程中,截面与轴之间保持固定的角度。截面可以是任何形状,在移动过程中它的尺寸可能是变化的,轴线也不一定是直线。

复杂的物体经常是由一些广义柱体相接在一起而组成的。考虑不同层次间表示法的相互影响,一种可能是信息从下向上流往各个表示法,并且每一层的计算只取决于前面的相邻层中产生的描述。例如,二维半简图的计算只需要初始简图中的信息,而不需要直接从图像中来的信息,也不利用任何关于从图像可以看到什么东西的线索。在计算中所用的每个过程独立地或近乎独立地进行。

另一种可能是,在初始简图和二维半简图中的各种计算是靠混杂的约束传播来进行的。例如,从三维处理得到的信息可以用于改进根据灰度的影调变化对表面方向的计算,反之亦然。信息可以在所有方向上流动。

还有一种可能是信息自顶向下流动,对图像的理解在很大程度上取决于有控制的想像。这样,早期视觉就由关于应该看到什么的固定的预测所引导。

## 5.5 拓扑视觉理论

视知觉研究200多年的历史,始终贯穿着“原子论”和“整体论”之争。原子论认为,知觉过程开始于对物体的特征性质或简单组成部分的分析,是从局部性质到大范围性质。而整

体论却认为，知觉过程开始于物体的整体性的知觉，是从大范围性质到局部性质。

前面讨论了马尔建立的视觉信息处理的计算的总体框架。在初始简图层次，视觉信息主要是两种，即强度的改变和图像的局部几何性质。这些局部几何性质是指平行性、相对位置和朝向等。视觉的最初过程就是计算这些局部特征或性质的表示，这些表示不外乎是图像的某些简单的和小的部分，如线段、柱体，以及它们之间的局部关系。然后，以这些局部性质的计算结果为基础，视觉信息处理进入更高水平的层次，如二维半和三维的计算。“什么是视觉信息最粗略的表示”？这个问题是视觉研究的一个中心问题，是建立任何一种视觉图像识别的模型都必须首先回答的问题。视觉系统对外界图像的知觉是从何做起的，对这类问题的回答是什么，将决定整个图像识别模型的方向。马尔的初始简图对这个问题的回答是，图形的简单部分和它们的局部性质。因此，马尔的视觉计算体系是从局部性质出发的，是由局部细节性质的计算到整体大范围性质的计算。

1982 年陈霖在《科学》杂志上就知觉过程从哪里开始的根本问题，原创性地提出了“拓扑性质初期知觉”的假说[133]。这是他在视知觉研究领域的独创性贡献，向半个世纪以来占统治地位的初期特征分析理论提出了挑战。在三十多年的时间里，在与持不同学术观点的学者的争论与交流中，他用令人信服的实验不断完善和论证着这一假说，使之成为被越来越多的国际同行所接受的学说，进而成为有国际影响力的理论。与传统的初期特征分析理论根本不同，拓扑性质初期知觉理论从大范围性质到局部性质的不变性知觉的角度，为理解知觉信息基本表达的问题，为理解知觉和认知过程的局部和整体的关系问题，为理解认知科学的理论基础——认知和计算的关系问题，提出了一个理论框架。

一系列视知觉实验表明，视图形知觉有一个功能层次，视觉系统不仅能检测大范围的拓扑性质，而且较之局部几何性质视觉系统更敏感于大范围的拓扑性质，对由空间相邻关系决定的大范围拓扑性质的检测是发生在视觉时间过程的最初阶段[853]。

拓扑学研究的是在拓扑变换下图形保持不变的性质和关系，这种性质和关系就称为拓扑性质。所谓拓扑变换是一对一的连接变换，它可以形象地想像成橡皮薄膜的任意变形，只要不把薄膜剪开或不把薄膜的任意两点粘合起来。一张橡皮薄膜可以任意地变形，可以从一个三角形变成一个正方形，三角形可以变成圆形或任意不规则的图形（图 5.15），只要不把它剪开。作为一个连通的整体这个性质，即连通性，仍然保持不变的。所以连通性是一种拓扑性质。另外，一个连通的图形中有没有洞或者有几个洞，这种性质也是一种典型的拓扑性质。

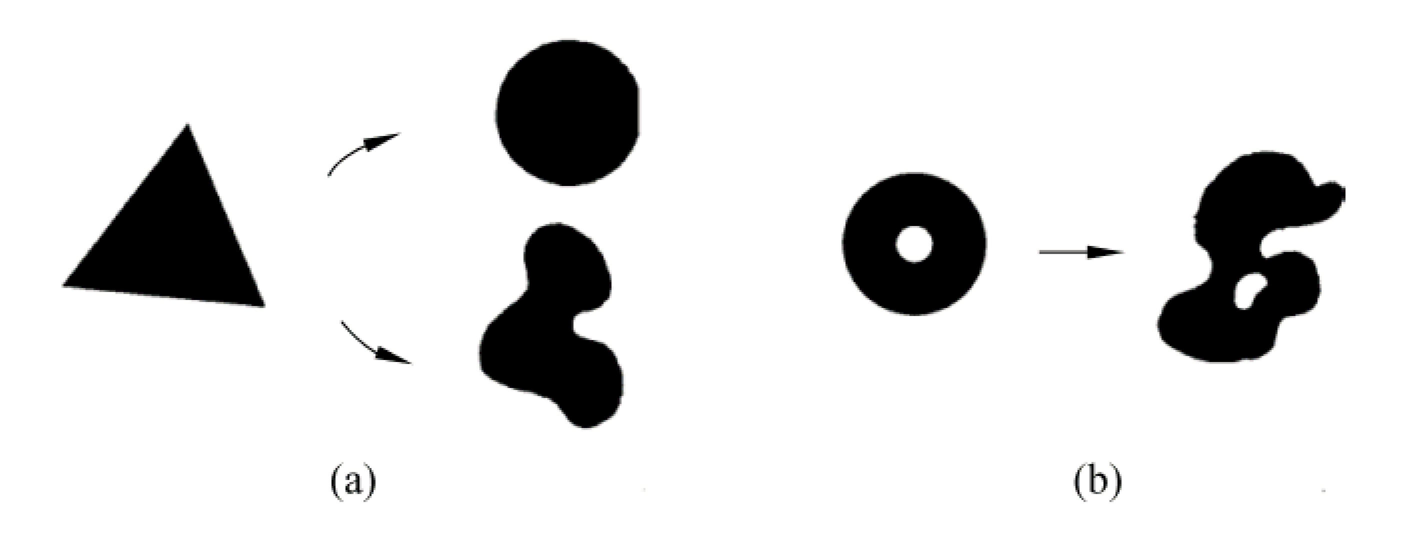



图 5.15　拓扑变换和拓扑性质的图示

人们直觉的经验,圆、三角形和正方形看起来是很不相同的图形,但是从拓扑学的角度来看,它们都是拓扑等价的、相同的。而圆和环,由于一个含有一个洞,另一个不含有洞,它们是拓扑不同的。尽管在通常的视觉观察的条件下,从人们在心理学相似性的角度来说,人们会觉得圆和环比较比圆和三角形、正方形要相像一些,但是如果视觉系统具有初期提取拓扑性质的功能,那么我们应当预计,在不能把圆和三角形、正方形区别开来的短暂呈现的条件下却仍然有可能把圆和环区别开来。图 5.16 表示用于这类实验的三组刺激图形。它们分别是实心圆和实心正方形、实心三角形、环。受实验者被要求注视每幅图的中心的黑点,然后每幅图被呈现短暂的 5ms,并且在撤去之后立即呈现另一幅空白的没有图形的蔽掩刺激,来干扰视觉系统对在此以前呈现的图形的知觉。受实验者被要求回答的问题并不是被呈现的在注视点两旁的图形是什么样的图形,而是被呈现的两个图形是一样的或是不一样的。

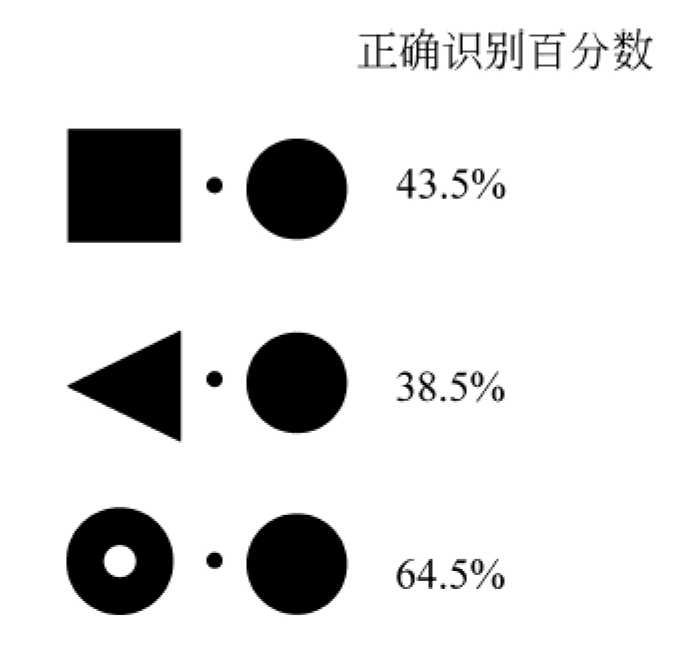


图 5.16　视觉系统对拓扑差异的敏感性

实验的结果也表示在图 5.16 中[853]。主要的实验发现是,视觉系统确实更敏感于拓扑性质的差异,也就是敏感于具有一个洞的环和没有洞的实心圆的差别。对圆和环一组刺激图形的正确报告率,要显著高于圆和三角形的正确报告率与圆和正方形的正确报告率。而且,拓扑性质等价的两对图形,圆和三角形与圆和正方形,它们的正确报告率的区别却没有达到统计意义,从而作为对照实验加强了视觉系统对圆和环的差别的敏感就是对它们之间的拓扑差异敏感的假设。这个同日常经验不一致却跟拓扑学的解释一致的实验,提供了一个支持拓扑结构假设的较为直接和令人信服的证据。

陈霖于 1983 年采用了称为“运动竞争的技术”,进行了似运动的实验。两幅刺激图形先后呈现。在第一幅中包含位于正中的单个刺激图形,而第二幅包含位于两边的两个跟中心等距离的刺激图形。对每个受实验者,调整两幅刺激图形的各自的呈现时间和相互的呈现时间间隔,使得受实验者在这样的呈现条件下产生似运动的知觉。受实验者回答的问题是,中间的包含在第一幅刺激呈现中的图形看起来是向包含在第二幅刺激呈现中的哪一个图形运动,简而言之,就是中间的那个图形看起来是向左边的图形运动,还是朝右边的图形运动。上述的运动竞争的实验方法给我们提供了一个客观量度来研究各种图形结构性质对似运动的影响。

如图 5.17 所示的 7 对刺激图形代表了各种拓扑不变性质的比较。每对刺激图形的中间一个图形包含在第一幅刺激呈现中,两边的两个图形包含在跟随第一幅呈现的第二幅刺激呈现中。为了帮助读者容易掌握这些刺激图形设计的规则,在图 5.9 中每个中间的图形都安排成跟右边的图形拓扑等价或具有相同的简单连通拓扑不变性质,而跟左边的图形拓扑不等价。当然,在实际的实验中,第二幅刺激呈现时,两个图形的任一个是随机地呈现在左边或右边,以避免受实验者的心理倾向性影响实验的正常结果。下边,我们仅选取这七对刺激图形的一部分作进一步说明。

第一对的三个图形每一个都是由完全相同的三条线段构成的。中间的图形 a 是一个箭头状的图形。而左边的图边 b 和右边的图形 c 都是由中间的图形改变而来,它们之间虽然

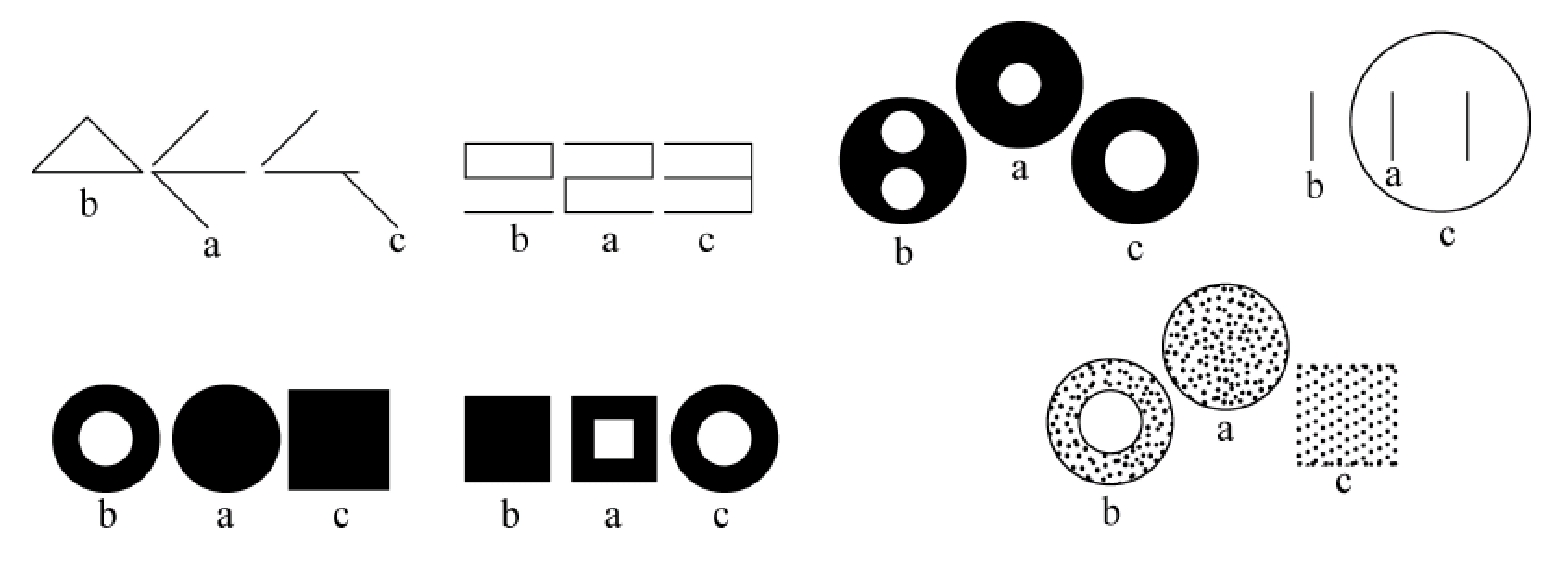



图 5.17　似运动刺激图形

有一个是封闭的而另一个不是封闭的拓扑性质的不同，但是它们都是把图形 a 的一条线段平移相同的距离而得来的。乌尔蒙(Ullman)解释似运动机制的一个代表性的计算理论，线段是建立两个似运动的图形之间一致性对应关系的基本特征，那么图形 b 和 c 较之图形 a 有相同的一致性对应关系，图形 a 不应当有向图形 c 方向运动的优越性。这样，这一对刺激图形给我们提供了一个机会，验证在似运动产生的机制中究竟是拓扑性质的检测还是从局部特征出发的计算在起作用。

运动知觉领域中，各种行为实验结果所得到的一个普遍接受的结论是，似运动和真实运动是等价的。如果这种观点正确，那么大脑背侧通路(Dorsal Pathway)的 MT 区就很可能是与似动相关的区域，因为大量的包括来自 PET 和 fMRI 的实验证据都表明 MT 区是专司运动的，不论是真实运动还是运动错觉(如运动后效)。另一方面，大量的研究还表明，形状知觉是由腹侧通路(Ventral Pathway)负责的。我们能看见两个不同形状的图形(如圆和方形)之间仍然产生似运动，因而可能会认为形状可能对似运动知觉不那么重要。卓彦等人 2003 年发表在 Science 上的文章[843]，带来了令人吃惊的结果：腹侧位置上的颞下回和颞中回前部有显著激活。而公认的运动区一背侧 MT 区却没有被激活。

2005 年，陈霖在 *Visual Cognition* 第四期上发表长达 88 页的"重大主题论文"[136]，对拓扑视觉理论概括为："知觉组织的拓扑学研究基于一个核心思想和包括两个方面。核心思想是，知觉组织应该从变换(Transsformation)和变换中的不变性(Invariance)知觉的角度来理解。两个方面是，第一方面强调形状知觉中的拓扑结构，这就是，知觉组织的大范围性质能够用拓扑不变性来描述；第二方面进一步强调早期拓扑性质知觉，这就是，拓扑性质知觉优先于局部特征性质的知觉。"优先"有两个严格的含义：第一，由拓扑性质决定的整体组织是知觉局部几何性质的基础；第二，基于物理连通性的拓扑性质知觉先于局部几何性质的知觉。

## 5.6　视觉的正则化理论

初级视觉是由一组能从二维光强度阵列恢复三维可见表面物理性质的处理过程组成的，它们的组合输出大体上对应于马尔的二维半图。初级视觉模块的计算理论主要解决表

象和处理这两个密切相关的问题,该理论必须说明输入和所期望的输出应该具有什么样的形式或表象,并提供把一种表象变换成另一种表象的算法(处理过程)。

初级视觉的精确定义是:它是光学成象问题的逆问题。经典光学或计算机图像学的基本问题是要确定三维物体的图像,而视觉则面临着从图像重建表面这样一个逆问题。由于在把三维世界投影成二维图像时损失了大量信息,因此视觉时时都要依靠自然的约束即关于客观世界的假设,才能获得确定的输出,找出并使用这些约束条件,这是在分析特殊视觉问题时会反复出现的问题。

运动的计算和图像强度变化的检测(物理边缘的检测)是初级视觉中两个重要的问题,它们充分地体现了初级视觉问题的困难所在。测量局部运动只能提供速度矢量的法向分量,切向分量仍然是"不可知的",因此估计整个速度场这个问题,一般说来,单靠从图像中直接可获得的测量是无法完全确定的,视觉流的测量本质上是不确定的。边缘检测所遇到的问题略有不同,边缘检测指的是这样一种处理过程,它要从图像的强度变化找出三维表面的物理边界。其实这个问题就是对图像数据进行数值微分,但是在成像和采样过程中所不可避免的噪声却使微分运算处于困难的境地,微分放大了噪声,因此这个处理过程本质上是不稳定的。

波杰奥(T Poggio)等指出,一个问题是适定的,如果它的解是存在的、唯一的、且连续地依赖于初始数据[586]。不适定问题则无法满足上述判据中的某一条或某几条。从形式上不难证明,初级视觉中的一些问题,诸如体视匹配、从运动恢复结构、视觉流计算、边缘检测、从明暗恢复形状、光度计算、表面重建等,都是在哈达玛德(Hadamard)意义下的不适定问题。视觉流计算之所以是不适定的,是因为从一条轮廓线上的法向速度分量恢复整个速度矢量场这个"逆"问题不能满足唯一性条件。边缘检测即数值微分之所以也是不适定的,则是因为解并不连续地依赖于数据。求解不适定问题即恢复"适定性"的主要想法,是引入适当的先验知识来限制容许解的集合。我们把任何一种使不适定问题适定化的方法统称为正则化。变分正则化是指在某一变分原理下重新表述一个不适定问题的正则化方法。如从"数据"$y$ 中找到 $z$ 这样的不适定问题。

$$Az = y \tag{5.7}$$

正则化,就需要选择范数 $\| \cdot \|$,并选择一个起稳定作用的泛函 $\| Pz \|$。在标准正则化理论中,A 是一个线性算子,范数是二次的,$P$ 是线性的。可供使用的两种方法是:

(1) 在满足 $\| Az-y \| < \varepsilon$ 的 $z$ 中找出使

$$\| Pz \|^2 \tag{5.8}$$

最小化的 $z$。

(2) 找出使

$$\| Az - y \|^2 + \lambda \| Pz \|^2 \tag{5.9}$$

最小化的 $z$,式中 $\lambda$ 是所谓的正则化参数。

迄今在初级视觉中使用的大多数稳定性泛函都是蒂克霍诺(Tikhonov)型的,也就是期望解 $z$ 的前 $P$ 阶导数的线性组合。表 5.2 列出了已经用变分原理解决了的几个初级视觉问题。

表 5.2 初级视觉的正则化

| 初级视觉问题 | 正则化原理 |
|---|---|
| 边缘检测 | $[(Sf-i)^2+\lambda(fxx)^2]\mathrm{d}x$ |
| 视觉流(基于面积) | $[(i_x u+i_y u+i_i)^2+\lambda(u_x^2+u_y^2+v_x^3+v_y^2)]\mathrm{d}x\mathrm{d}y$ |
| 视觉流(基于轮廓线) | $[(V\sharp N-Nn)^2+\lambda(\partial/\partial s)V^2]\mathrm{d}s$ |
| 表面重建 | $[(s\cdot f-\mathrm{d})^2+\lambda(f^2xx+2f^2xy+f^2yy)]\mathrm{d}x\mathrm{d}y$ |
| 时空逼近 | $[(s\cdot f-i)^2+\lambda(\nabla f\sharp V+ft)^2]\mathrm{d}x\mathrm{d}y\mathrm{d}t$ |
| 颜色 | $\Vert I^v-Az\Vert^2+\lambda\Vert Pz\Vert^2$ |
| 恢复形状 | $[E-R(f,g)^2+\lambda(f_x^2+f_y^2+g_x^2+g_y^2)]\mathrm{d}x\mathrm{d}y$ |
| 体视 | $\{[\nabla^2 G(L(x,y)-R(x+\mathrm{d}(x,y),y))]^2+\lambda(\nabla\mathrm{d})^2\}\mathrm{d}x\mathrm{d}y$ |

根据 Horn 和 Schunk 的工作,希尔德莱斯(Hildreth)对速度场提出了更一般的平滑性约束,在物理上这个约束基于这样的考虑,即真实世界是由表面平滑的固体组成的,而这些固体的投影速度场一般说来是平滑的。第一种算法是在测量速度矢量的法向分量之后,假设法向速度分量 $V^N(s)$ 是精确测定的情况下,使

$$\Vert PV\Vert^2=\int\left(\frac{\partial V}{\partial s}\right)^2\mathrm{d}s \tag{5.10}$$

达到极小(其中 $s$ 为弧长)。积分是沿轮廓线进行的。对不精确的数据则可采用第二种方法,由

$$\Vert V\cdot N-V^N\Vert^2+\lambda\int\left(\frac{\partial V}{\partial s}\right)^2\mathrm{d}s \tag{5.11}$$

的极小化给出所求的解,式中 $N$ 为轮廓线法向单位矢量,$\lambda-1$ 表示数据的可靠程度。

对于正则化分析来说,头等大事是解在物理上的合理性,而不是解的唯一性。对一个问题以及对问题的主要约束所进行的物理分析,在正则化中起着重要的作用。在某种特殊情况下,正则解并不对应于具有实际物理意义的解,这时,为求解不适定问题而作的先验假设就可能被违背。在一些很一般的条件下、平滑性假设可以给出正确的解。但是对于某几类运动和轮廓线,平滑性原理却无法产生正确的速度矢量场。人的视觉系统在这种环境中似乎也只能得出同样是不正确的速度矢量场,因此,这种现象有助于我们揭示大脑对外部世界所作的先验假设。

生物视觉中一个令人百思不得其解的谜就是它的高速度,对于这个问题人们总是提出平行处理作为它的理由。然而,由数字处理得到的计算模型却是不能令人满意的,特别是有越来越多的证据表明,神经不是一个复杂的装置,远不同于简单的数字开关。因此,研究一下初级视觉的正则化方法是否可以导致一种不同类型的并行算法,这是很有意义的。我们已经指出,一个线性的模拟网络(电的或化学的)乃是求解标准正则化理论所要求的变分原理的一条自然途径。

从变分原理到电学网络或化学网络这种映射的基本出发点是哈密顿最小作用量原理。可以用模拟网络进行计算的变分原理的类型由基尔霍夫电流、电压定律给出,这两条定律代表每一个网络元件所满足的保守性约束和连续性约束。一般说来,实现一个变分原理的网络可能不止一个,而是有许多个。例如,霍普菲尔德在解决联想记忆问题时所提出的那种级联网络也可用于标准正则化原理。

根据基尔霍夫定律可以证明,对每一个存在唯一解的二次变分问题,都存在一个与之相应的、具有同一解的、由电阻和电压源或电流源构成的电学网络。从当前对神经元、膜和突触的生物物理学性质的了解来看,这种模拟的并行计算模型就特别有意义。一小片神经元膜就可等价于电阻、电容和现象学意义下的电感。树突尖端处的突触模拟了电压源,而大树突或胞体上的突触则模拟了电流源。因此,单个神经元或神经元的小网络就可实现正规化原理下的模拟解。

初级视觉的这一新理论框架清楚地表明了标准的蒂克霍诺(Tikhonov)型正规化理论的内在优点和局限性。它的主要问题是:需要了解待求未知函数的平滑化程度。标准正规化理论只涉及线性问题,其基础是二次型稳定器。由此得到的结果是二次泛函最小化和线性的欧拉—拉格朗日方程。为了加上正确的物理约束,可能需要非二次泛函。即使在这种情况下,标准正则化理论也还是可以使用的,但解空间不再是凸的,因而在极小化过程中可能会发现许多局部极小。有人已经提出了一个从深度数据重建表面时可以保持不连续点的非二次稳定器。

视觉正则化理论现在所面临的挑战,是如何使它突破标准正则化方法的限制。可以用二次泛函来实现的计算,其适应性是很有限的。只要认识到对二次代价泛函的极小化只能产生一个线性正则化算子,也就是说,只能产生从输入数据到解空间的一个线性映射,我们就可以看清这一点。在数据都在规则的格点上且满足适当的条件这个特殊情况下,这个线性算子就可以变成一个卷积运算,也就是对数据作简单的滤波处理。与物理学中的线性模型一样,标准正则化理论在许多情况下是一种非常有用的近似方法,但它并不能解决视觉中的全部复杂问题。

实现正则化的另一个严格的方法是在贝叶斯估计和马尔科夫随机场模型的基础上建立起来的。在这种方法中先验知识是用恰当的概率分布来表达的;而在标准正则化方法中,先验知识则产生对解空间的约束。在希尔德莱斯的运动计算中,平滑化假设相当于假设在一条轮廓线上两个相邻点之间的速度矢量变化是零均值的、独立的高斯随机变量。随机方法与标准正则化方法之间的这一联系,使我们得以洞察约束特性和稳定器选择的真相。在解决视觉这个求逆问题中所使用的变分原理,就对应于能产生合理解集的马尔科夫结构。

在今后的研究中,与视觉正则化相关的一个领域是正则化算子题。在标准正则化条件下,把数据影射到解空间的相应的线性算子可以通过一种联想学习方法获得。这种学习方法是在研究生物记忆问题时提出来的。

## 5.7　基于模型的视觉理论

人类视觉最显著的特点之一是有选择性。这是指观察者的注意力总是有目的地指向他最感兴趣的事物。一般生物最注意的是环境中时常变化的事物,忽略固定不变的事物。因为这样就可以迅速辨别出什么是对自己有益的,什么是对自己有害的。从而作出攫取或躲避反应。另一个重要的特点,如格式塔心理学家发现的那样,是人类具有对图像数据进行组织归纳的能力,也就是在多个层次上发现图像数据的规则性(Regularity)、一致性(Coherence)、连续性(Continuity)等整体特性的能力。实验证明,人类视觉系统具有在低层处理中获取图像拓扑特性的能力。

马尔理论完全不考虑视觉中的选择性和整体性，把初级视觉研究的目标确定为按照各种物理模型和附加约束条件，根据图像中各点灰度或其他测量结果，恢复景物中表面的有关特性，如表面方向、深度、反射率等。但由于图像中各点的灰度是光照，表面材料的反射特性、表面方向、观察方位等多种因素共同作用的结果。并且在成像过程中失去了各点的距离信息，所以，根据图像中的测量值（如灰度）恢复相应表面的三维特性（如，深度、方向），从本质上来说是一个约束不充分（Underconstrained）的问题。也就是说，图像的测量值本身不能提供充分的信息来恢复相应表面的三维信息。因此，为能根据马尔理论恢复表面的三维信息必须增加附加的约束条件。例如，把物体仅限于刚体的范围，假设表面是连续的，各向同性的；或更为特殊的约束，如表面是由平面构成，点光源照明，材料的反射率为常数等。这些约束条件只能在某些人造环境下（例如在所谓的“积木世界”）得到满足，而在自然界或实际情况下通常是不能满足的。而且即使具备了这些条件，目前采用的大多数求解方法类似于求解经典的边值问题。总的来说性能比较脆弱，容易出错。马尔理论的这些困难在 20 世纪 80 年代末已经暴露得较为明显。

由上述分析可知，现有的两种视觉信息的处理理论各自遇到了严重的困难，还都不能自成系统地、可靠地处理视觉问题。因此，有的研究者提出了各种设想对上述理论提出了修改，并试图把这两种方法以取长补短的方式结合起来。

基于模型的视觉理论认为信息的概念是与从一组候选的对象中作出选择相联系。如果不知道一组可供选择的刺激或响应，人们就对刺激或响应无从说起。此外，人们还必须知道定义这组候选物和对这组候选物的成员进行区别的特性或特征。而且随着要解决的任务不同，这些特征和特性也不同。例如，在视觉敏感中，刺激引起在视网膜的一组可能的状态中选择一种状态，并得到一幅图像。在感知中，选择是根据不变量（Constancies）和参数（Parameters）做出的。如果一个婴儿能听到声音，但他的感知不变量只包括“安静”和“噪声”的话，那么任何音乐对他来说都将包括一样多的信息，而这些音乐对一个训练有素的音乐家来说就会包含丰富得多的信息。

这种视觉理论利用特征检测器的概念作为把点状的图像数据与宏观信息相联系的桥梁[556]。因此，基于模型的视觉理论体现了格式塔理论中的选择型和整体性。

有的系统利用模型来帮助对不完全线画图进行解释。它先分析一幅线画图，并对图中的物体种类和方位作出假设。然后预测所假设景物的线画图，并最后试图校验所假设的线画图。这种主动地应用模型的方法被称为基于模型的、自顶向下的、语义的和目标驱动的方法。在某些系统中模型只是简单地校验结果，而在其他一些系统中则完全由模型来控制在何处可见到何物。在 20 世纪 70 年代的初期，学术界普遍认为不引入自顶向下的景物知识的初级视觉处理从本质上看无法产生丰富的、有用的描述，并认为一个聪明的视觉系统应该对景物中有些什么事物了解很多。这类方法中有代表性的程序有希雷（Shirai）的语义边缘检测方法、亚克姆斯基（Yakimosky）的基于语义的区域分析方法，和特南鲍姆（Tenenbaum）的解释导引的分割方法等。但这些方法在高级语义知识如何与低级的聚集处理相衔接上遇到严重的困难，并且不可避免地要引入一些过于简单的假设，因此并未取得预期的成功。这样，在 20 世纪 70 年代后期计算机视觉研究面临左右为难的困难境地。一方面认为没有关于图像是什么物体的先验知识，就难以达到完善的图像分割，也就难以理解图像；另一方面在应用先验知识来导引低层处理进行图像分割时又遇到了上述的严重困难，于是很自然地

提出一个问题：在不知道图像中有些什么物体的条件下，二维的灰度图像到底能提供什么样的信息。例如，人是通过用双目观察事物来获取深度信息的，那么人是在识别出看到的是什么物体以前就获取了深度信息呢？还是不论是否知道看到的是什么物体都能获得深度信息？再例如，灰度图像能否提供物体的形状信息等。对这些问题的回答，不可避免地要涉及对人类视觉的研究。

## 5.8 计算机视觉

计算机视觉是计算机科学和智能科学的重要组成部分，研究人类视觉的计算模型，利用计算机对描述景物的图像数据进行处理，以实现类似于人的视觉感知功能。

人的视觉能力很强，似乎毫不费力就可以对周围的景物一目了然。人们研究要使计算机具有或接近人的视觉能力，必须对人类自身的视觉感知机制十分了解。因此，智能科学的任务之一就是要揭示人类视觉的机理。

通过对图像或图像序列的分析，得到景物的尽可能完全和正确的描述。研究完成这一任务的理论、方法及其软硬件实现，称为图像理解，许多人把它看成是计算机视觉的同义词。由于物体的几何形状、物理性质、摄像机特性、照明情况以及物体与摄像机之间的空间关系等诸多因素在成像过程中被综合，仅仅表现为图像中像元的灰度值，以解决其反问题为目标的图像理解或计算机视觉，无疑是十分困难的，它的研究涉及到神经科学、应用数学、图像处理、模式识别、知识工程等多个学科分支的相互交叉和渗透。但从应用的观点来看，却有许多办法可以由图像中提取信息来解决与景物有关的实际问题，而不需要完全确定景物的三维结构。研究这种面向应用的计算机视觉系统的设计与实现技术，称为机器视觉。

计算机视觉的研究已有三十多年历史。早期研究工作集中在二维的情形，例如字符识别，染色体显微图像分类等。1965年，罗伯特(L G Robert)设计和实现了一个理解多面体积木世界景物的程序，是计算机三维视觉的最早成果。在该程序中，建立了一个含有立方体、长方体、楔块和六棱柱的三维物体模型库。由线画图识别这些物体的方法是先根据顶点数目之类的简单特征确定候选的三维模型，再经过平移、旋转、尺度变换和投影，与待识线画图进行匹配。20世纪70年代中，计算机视觉在遥感图片解释、生物医学图像分析、工业自动检验等方面都取得了不同程度的成功。随着传感技术的发展，开始了距离图像获取和分析的研究，如由移动传感器对静止景物或静止传感器对运动物体获取的图像序列的分析，已成为受到重视的新方向；这种通过对运动的分析，有可能确定物体的三维形状和运动参数，可用于移动机器人导航或工业机器人视觉系统。1978年，巴卢(H G Barrow)和特南鲍姆提出可以通过产生各种固有的景物描述性表示如深度信息、表面朝向等来改进机器视觉系统的处理能力。20世纪80年代初期马尔在此基础上建立了比较系统和完整的视觉计算理论。马尔理论开创了计算机视觉蓬勃发展的新局面，并主导了整个20世纪80年代的研究工作。直到1990年，阿罗莫讷斯(J Aloimonos)提出有目的的、定性的主动视觉等新理论框架，才形成当今计算机视觉研究的新热点。

计算机视觉方法的研究经历了几个发展阶段。在20世纪80年代以前，多数研究工作都是针对个别的有限制的问题进行的，其基本方法是：①获取灰度图像；②从图像中提取边缘、周长、惯性矩等等特征；③从描述已知物体的特征库中选择特征匹配最好的结果。由

于缺乏三维信息，上述方法很难在复杂的环节中取得较好的效果。

20 世纪 80 年代是计算机视觉方法的一个新发展阶段，按照马尔理论（参见视觉计算理论），视觉是通过自下而上的三个层次的信息处理过程来实现的，即

（1）对图像进行边缘检测与图像分割等低层视觉处理。

（2）求取深度信息、表面朝向等 2.5 维描述，主要方法有由影调、轮廓、纹理等恢复三维形态，由体视恢复景物的深度信息，由图像序列分析确定物体的三维形状和运动参数，距离图像获取与分析以及结构光方法等。

（3）根据三维信息对物体进行建模、表示与识别，可采用基于广义圆柱体的方法（参见视觉计算理论），另一常用方法是将物体外形表示为平面或曲面块（简称面基元）的集合，每个面基元的参数以及面基元之间的相互关系用属性关系结构来表示，从而将物体识别问题转化为属性关系结构的匹配问题。

到 20 世纪 80 年代末，尽管在上述三个层次信息处理模块的计算理论、算法与表达乃至硬件实施都有很大进展，仍然不能实现集成起来的通用计算机视觉系统。于是，一些学者开始寻找突破马尔理论框架的新途径。以 1990 年阿罗莫讷斯（J Aloimonos）提出定性视觉、主动视觉等为标志，计算机视觉方法发展到当前的新阶段。定性视觉方法的核心是将视觉系统看成执行某一任务的更大系统的子系统，视觉系统所要获取的信息，只是完成大系统任务所必需的信息。

在许多情况下，只要景物的定性或不完全描述就够了，无须如马尔理论所要求的完整的定量描述。景物的描述与识别，也不只经历一个自下而上的单向信息处理过程，而是在自下而上与自上而下相结合的控制策略下实现的。主动视觉方法则集感知、规划与控制为一体，通过这些模块的动态调用和信息获取过程与处理过程的相互作用，来更有效地完成视觉任务。该方法的核心是主动感知机制的建立，就是根据当前任务、环境状况、阶段处理结果和有关知识，来规划和控制下一步获取信息的传感器类型及其位姿。实现多视点或多传感器的数据融合，也是其关键技术。

计算机视觉的发展，提高了自动化和机器智能水平，为发展智能机器人和各种智能系统，提供了一项关键技术。随着主动视觉、定性视觉、多传感器信息融合等新方向的深入研究，以及更多地面向应用，可以预料，计算机视觉将有更诱人的发展。

### 5.8.1　图像分割

图像分割是将一幅图像分成不同的部分或区域的过程。图像分割是在数字图像处理、图像分析及图像理解中的一个重要的基本步骤。其目的是把图像分成一些有用的或有意义的部分或区域，以便进一步对图像进行分析与理解。把图像分割成有明确含义的（代表各种物体或背景）区域的方法：一种是根据各图像像素之间的不同灰度或不同的颜色分量来进行分割。这称之为基于像素的分割方法。另一种是基于不同类型的区域在图像中的不连续性来进行分割，称之为基于区域的分割方法。例如，我们可以根据不同区域有不同的纹理（组织结构特征）来进行分割。此外，图像中的不连续也形成了边界，我们也可以利用边界来进行图像分割。基于像素的方法可以用来分割白纸黑字，或黄色的墙和棕色的门。但是，要分割一幅图像中的两种不同的花布，就要求助于基于区域（纹理组织）的方法了。基于像素的分割方法还可以再分为利用门限或直方图的方法（参见图像像素分类）。基于区域的分割

方法还可以再分为利用边界(边缘)、利用区域以及同时利用边界和区域的方法。

图像中的边缘是图像中两个区域的边界,而这两个区域在亮度上(或颜色上)有相当差别。通常认为,边界处的亮度有较突然的变化。从数学的角度看,边界处的亮度对空间(二维图像平面坐标)的导数有较大的绝对值。所以,大部分的边缘检测技术都基于求导数的方法。由于计算机并不是用连续的函数处理大部分图像,而是用一些离散点上的值来表示。所以求导数的算法也必须用离散的差分算法(在图像处理中,差分是用一些提取特征的算子来表示的,可参见图像特征提取中有关内容)。在边缘检测中,经常用到亮度梯度的概念。图像亮度梯度是一个向量,其分量是图像亮度沿某两个坐标方向求偏导数所获得的值。图像 $f(x,y)$ 看成一个连续函数可以表示为

$$G=\begin{bmatrix}G_x\\G_y\end{bmatrix},\quad G_x=\frac{\partial f(x,y)}{\partial x},\quad G_y=\frac{\partial f(x,y)}{\partial y} \tag{5.12}$$

在实际的数字图像上,梯度的两个分量可以用下面的模板算子来计算

$$\begin{bmatrix}-1 & 1\\-1 & 1\end{bmatrix} \quad 和 \quad \begin{bmatrix}1 & 1\\-1 & -1\end{bmatrix}$$

梯度的幅度是

$$\sqrt{G_x^2+G_y^2} \tag{5.13}$$

梯度的方向是

$$\tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{5.14}$$

上面提及的算法以及在图像特征提取中的算法都是利用较小邻域的信息对孤立点来进行的。通常在采用这些算法之后,还要作进一步的检测及连接工作,以获得有意义的物体的边界或者物体表面的分界线。为了进一步确认及连接边界,一般可采用两种方法:局部分析法和全局分析法。

(1) 局部分析法是最简单的一种确定连接的方法。考虑一个小的邻域,例如 3×3 或 5×5 的邻域,在这个邻域里,我们把相似的点连接起来形成具有某种共同的特性的边界。通常,相似性要考虑两方面:一是用边界检测算子(梯度)计算后获得的强度(梯度的幅度);二是导数值沿两个坐标方向的比例(梯度的方向)。在某个图像区域内,如果检测出来的两点在梯度幅度及方向上的差别都小于某一给定值,我们就认为它们属于同一条边界,可以在图像中把它们标记出来。这个过程可以对整个图像分块地重复进行。

(2) 全局分析法,采用全局分析法时,要研究被检测出来的可能是边界点的像素在整个图像平面中的相互关系。霍夫变换(参见图像特征提取)可以用来进行全局性的分析确定。

边缘检测法通过在各个像素周围窗口中进行的局部运算可确定这个像素是否在边缘上。通常所作的运算是微分或与理想的边缘作相关运算。然后对所得结果再作阈值运算,以产生二值图像。被检测到的边缘点要连接成有意义的线段。这可通过按某种规则对边缘点进行跟踪来完成。在跟踪过程中会遇到边缘之间的间隙,曲线的突然变化或由于噪声引起的假边缘点等情况。这将要求复杂的跟踪规则。

区域分割法通过寻找其中的像素具有相同的灰度(或特性)的区域来完成图像的区域分割。这两种方法是互补的,在理想情况下得到的结果是等价的。图像区域分割是基于图像区域特性的差异对图像进行分割的技术。区域分割的基本思想是标识图像中各个具有相似

特征的区域。相似的特征可以是形状、像素值或纹理等。在模式识别中的聚类技术也可用于基于区域的图像分割。

(1) 模板匹配是将图像中的区域和一组给定的模板进行比较匹配,从而将符合模板的物体从图像的其他部分中分割出来,而剩余的图像则可根据需要再用其他方法分析。例如,模板匹配可用于分割图文混排的书稿。当文字用模板匹配的方法找出来以后,图形可再用其他方法进行分析。模板匹配的过程往往用相关或卷积计算来进行(参见图像处理的基本运算)。

(2) 纹理分割。当物体置于明显的纹理背景中或物体本身具有较强的纹理特征时,就需要利用基于纹理的区域分割方法。由于纹理是某种模式,或者说图案、花样、结构等的重复,所以不能用单个的像素的特性(灰度或颜色)来描述。当然也无法用基于像素的分类方法。由于纹理经常包含有大量的边缘,因此,除非滤去纹理,否则用边界跟踪的方法分割有丰富纹理的图像很难有好的效果。

纹理的描述与分类是分割的基础(参见图像特征提取)。当我们知道图像中有某种纹理存在时,可利用已知纹理的特征(如该纹理在频域中的描述或空间灰度关系矩阵)在图像中寻找。如果事先没有知识的话,可以采用基于区域的聚类方法进行纹理区域的分割。一种容易想到的办法是:把图像分成若干(小)块,计算每一块的纹理特征,根据特征差别的程度决定是否把小块合并。

(3) 区域聚类法一般可分为区域生长法及分裂合并法。区域生长的基本思路是:从满足检测准则的点或一块区域开始,在各个方向上"生长"物体。"生长"的依据是:同一类型区域的特征,如灰度、颜色及纹理特征等,相差不会太远。满足一定合并条件的邻域可以并入该区域。在生长过程中,合并条件可以调整。当再也找不到可合并的邻域时,生长停止。区域的分裂和合并,这个方法的基本思路是:首先将图像分为若干"初始"区域,然后再分裂或合并这些区域,逐步改进区域分割的指标,直到最后将图像分割为数量最少(或符合某一要求)的"基本一致"的区域为止。通常,"一致"性的标准可用特性的均方误差来量度。与基于边界的图像分割方法(参见图像边缘检测)相比,基于区域生长法和分裂合并法对噪声相对不敏感,但是计算复杂度较高。

## 5.8.2 图像理解

对于以点阵(光栅)形式表示的、由各种亮暗或彩色的点集所组成的一幅数字图像或一组活动图像序列,如果计算机能够通过图像分析识别手段"懂得"该图像所表达的"语义"内容,能回答人们提出的关于该图像的"语义"内容的问题(例如,画面上有没有人、有几个人、每个人在做些什么等等),就称计算机已经"理解"了该图像(或图像序列)。图像理解一般由低层次的去噪声处理、边缘提取、区域分割、特征提取、目标物提取,以及高层次推理分析等过程综合而成,也可以先从高层次的先验知识或模型假设开始,逐步向低层次处理进行"假设验证"。这个过程是人们希望用机器模拟或更多地取代人的高级视觉和推理功能而设定的研究方向,也被认为是比通常的图像分析识别更为高级的图像处理,因为一般认为"识别"基本等同于模式识别中的分类或辨认,而"理解"则需要反复推敲和比较之后抽象出更深层的内涵。但是,对于同一幅图像,不同的人并不见得都得出相同的理解结果。感兴趣的问题不同,看同一幅图时的注意点也就不同。另外,对同一事物的理解深度、层次一般都因人而

异,有时还有互相对立的结论。所以,所谓图像理解目前主要是一个意向性的命题,尚缺严格的定义或理论模型。在实际应用中,一般需要将对象和问题限制在较小、较明确的范围内,才可能取得比较接近预期目的的结果。

### 5.8.3 主动视觉

20 世纪 80 年代中期由美国的一些研究计算机视觉的学者提出的一种新的视觉感知原理和设计视觉系统的方法。其主要思想,是根据视觉感知任务的要求或外部刺激,系统主动地改变传感器(摄像机)的内部参数(焦点、焦距、光圈、聚散度等)及外部参数(位置、方向、光源条件等),以达到简化视觉计算,消除景物解释的歧义性,实现对真实世界鲁棒性、实时感知的目的。主动视觉的机理更接近人的视觉系统,例如人在观察时,要不断地转动眼球,甚至转动头部或移动身体的位置,以调整视线和注意点。传统的计算机视觉理论(以著名学者 D. Marr 为代表),把视觉过程看成一个信息处理的过程,侧重于研究如何从景物的二维图像,通过由低到高的分层表示和模块化的视觉计算,最终获得三维景物的精确几何描述和景物中物体性质的定量描述。由于三维景物的成像过程非常复杂,景物中物体的几何形状、表面反射特性,以及光源、物体与摄像机间的空间关系等诸多因素都将对二维图像中像元的灰度值产生影响;另外,成像过程中还会引入噪声、畸变以及遮挡等问题。因此作为成像过程的逆过程,从原始的二维图像获取三维景物精确的描述是非常困难的任务。目前研究的许多算法不仅计算冗长繁杂,对噪声的鲁棒性差,而且有时无法得到唯一解,最终结果的可靠性和精度都无法保证,致使计算机视觉的应用进展缓慢。为此自 20 世纪 80 年代中期以来,许多学者提出了各种新的计算机视觉理论框架和思想。与主动视觉类似的还有"定性"视觉(Qualitative Vision),"有目的"视觉(Purposive Vision),"有生命"视觉(Animate Vision)等。虽然使用的术语不同,但中心思想都是将视觉感知与其所参与的、与环境交互的动作紧密结合起来。即根据有关任务的要求和知识,"有目的"地、"主动"地控制图像获取过程,有选择地(在空间、时间及分辨率等方面)进行观察,为视觉处理和计算提供更多的约束,以简化计算,增加鲁棒性,消除歧义性,最终获得为完成该特定任务所需的"定性"的描述,而不需要对景物的完整的、定量的描述。这一思想并不是对传统的计算机视觉理论的全盘否定,区别仅在于表示和算法更贴近自然,使用的工具仍相同。但无论最终采用什么样的术语(目前的趋势是"主动视觉"较通用),这一思想将为计算机视觉研究,特别是视觉系统的应用开辟新的途径。主动视觉研究的初期偏重于研究具有类似人眼特性的摄像机瞄准系统(多分辨率、多自由度),以及有关跟踪、凝视控制等低层处理问题。随着用于主动视觉研究的实验设备的逐渐普及,主动视觉研究的内容将不断扩展。另一方面,作为与外部环境进行交互作用的机器人系统,则是主动视觉的主要应用领域。机器人可以看成是感知和动作集成的智能体(Agent),例如手眼系统,移动机器人系统等。因此主动视觉如何集成到机器人体系结构中将是主动视觉的重要研究内容之一。换句话说,主动视觉的发展与智能机器人的研究密不可分。

### 5.8.4 立体视觉

立体视觉是计算机视觉领域的一个重要课题,它的目的在于重构场景的三维几何信息。

立体视觉的研究具有重要的应用价值，其应用包括移动机器人的自主导航系统、航空及遥感测量、工业自动化系统等。一般而言，立体视觉的研究有如下三类方法。

(1) 直接利用测距器(如激光测距仪)获得程距信息，建立三维描述的方法。程距法根据已知的深度图，用数值逼近的方法重建表面信息，根据模型建立场景中的物体描述，实现图像理解功能。这是一种主动方式的立体视觉方法，其深度图是由测距器获得的，如结构光、激光测距器等其他主动传感技术。这类方法适用于严格控制下的环境，如工业自动化的应用方面。

(2) 仅利用一幅图像所提供的信息推断三维形状的方法。依据光学成像的透视原理及统计假设，根据场景中灰度变化导出物体轮廓及表面，由影到形，从而推断场景中的物体。线条图的理解就是这样的一个典型问题，曾经引起了普遍的重视而成为计算机视觉研究领域的一个焦点，由此产生了各种各样的线条标注法。这种方法的结果是定性的，不能确定位置等定量信息，该方法由于受到单一图像所能提供信息的局限性，存在难以克服的困难。

(3) 利用不同视点上的，也许是不同时间拍摄的，两幅或更多幅图像提供的信息重构三维结构的方法。利用多幅图像来恢复三维信息的方法，它是被动方式的。根据图像获取方式的区别又可以划分成普通立体视觉和通常所称的光流(Optical Flow)两大类。普通立体视觉研究的是由两台摄像机同时拍摄下的两幅图像，而光流法中研究的是单个摄像机沿任一轨道运动时顺序拍下的两幅或更多幅图像。前者可以看作后者的一个特例，它们具有相同的几何构形，研究方法具有共同点。双目立体视觉是它的一个特例。

立体视觉的研究由以下几部分组成。

(1) 图像获取。用作立体视觉研究的图像的获取方法是多种多样的，在时间、视点、方向上有很大的变动范围，直接受所应用领域的影响。立体视觉的研究主要集中在三个应用领域中，即自动测绘中的航空图片的解释，自主车的导引及避障，人类立体视觉的功能模拟。不同的应用领域涉及不同类的景物，就场景特征的区别来分，可以划分成两大类，一类是含有文明特征的景物，如建筑、道路等；另一类是含有自然特征的景物和表面，如山、水、平原及树木等。不同类的景物的图像处理方法大不相同，各有其特殊性。与图像获取相关的主要因素有场景领域、计时、时间照明和阴影、成像形态(包括特殊的遮盖)、分辨率、视野、摄像机的相对位置等。

(2) 摄像机建模。摄像机建模就是对立体摄像机组的重要的几何与物理特征的表示形式，它作为一个计算模型，根据对应点的视差信息，用于计算对应点所代表的空间点的位置。摄像机模型除了提供图像上对应点空间与实际场景空间之间的映射关系外，还可以用于约束寻找对应点时的搜索空间，从而降低匹配算法的复杂性，减小误匹配率。

(3) 特征抽取。几乎是同一灰度的，没有特征的区域是难以找到可靠匹配的，因而，绝大部分计算机视觉中的工作都包括某种形式的特征抽取过程，而且特征抽取的具体形式与匹配策略紧密相关。在立体视觉的研究中，特征抽取过程就是提取匹配基元的过程。

(4) 图像匹配。图像匹配是立体视觉系统的核心，是建立图像间的对应从而计算视差的过程，是极为重要的。

(5) 深度计算。立体视觉的关键在于图像匹配，一旦精确的对应点建立起来，距离的计算相对而言只是一个简单的三角计算而已。然而，深度计算过程也遇到了显著的困难，尤其是当对应点具有某种程度的非精确性或不可靠性时。粗略地说，距离计算的误差与匹配的

偏差成正比,而与摄像机组的基线长成反比。加大基线长可以减少误差,但是这又增大了视差范围和待匹配特征间的差别,从而使匹配问题复杂化了。为了解决这一问题出现了各种匹配策略,如由粗到精策略、松弛法等。

在很多情况下,匹配精度通常是一个像素。但是,实际上区域相关法和特征匹配法都可以获得更好的精度。区域相关法要达到半个像素的精度需要对相关面进行内插。尽管有些特征抽取方法可以得到比一个像素精度更好的特征,但这直接依赖于所使用的算子类型,不存在普遍可用的方法。

另一种提高精度的方法是采用一个像素精度的算法,但是利用多幅图像的匹配,通过多组匹配的统计平均结果获得较高精度的估计。每组匹配结果对于最后深度估计的贡献可以根据该匹配结果的可靠性或精度加权处理。

(6) 内插。在立体视觉的应用领域中,一般都需要一个稠密的深度图。基于特征匹配的算法得到的仅是一个稀疏而且分布并不均匀的深度图。在这种意义下,基于区域相关匹配的算法更适合于获得稠密的深度图,但是该方法在那些几乎没有信息(灰度均匀)的区域上的匹配往往不可靠。因此,两类方法都离不开某种意义的内插过程。最为直接的将稀疏深度图内插成稠密的深度图的方法是将稀疏深度图看作连续深度图的一个采样,用一般的内插方法(如样条逼近)来近似该连续深度图。当稀疏深度图足以反映深度的重要变化时,该方法可能是合适的。如起伏地貌的航空立体照片的处理中用这种方式的内插也许是比较合适的。但是这种方法在许多应用领域中,尤其是在有遮掩边界的图像的领域中,就不适用了。

格里姆森(Grimson)指出可匹配特征的遗漏程度反映了待内插表面变化程度的相应限度,在这种基础上,他提出了一个内插过程。换一个角度来看,根据单幅图像的“由影到形”的技术,用已经匹配上的特征来建立轮廓条件和光滑的交接表面可以确保内插的有效性。这些方法结合起来,可以使内插过程达到合乎要求的目标。内插的另一种途径是在已有的几何模型与稀疏深度图之间建立映射关系,这是模型匹配过程。一般而言,要进行模型匹配,预先应将稀疏深度图进行聚类,形成若干子集,各自相应于一种特殊结构。然后找每一类的最佳对应模型,该模型为这种特殊结构(物体)提供参数和内插函数。如 Gennery 用这种方法来发现立体对图片中的椭圆结构,Moravec 用于为自主车探测地面。

双目立体视觉(Binocular Stereo Vision)理论建立在对人类视觉系统研究的基础上,通过双目立体图像的处理,获取场景的三维信息,其结果表现为深度图,再经过进一步处理就可得到三维空间中的景物,实现二维图像到三维空间的重构。Marr、Poggio、Grimson 最早提出并实现了一种基于人类视觉系统的计算视觉模型及算法。双目立体视觉系统中,获取深度信息的方法比其他方式(如由影到形方法)较为直接,它是被动方式的,因而较主动方式(如程距法)适用面宽,这是它的突出特点。

双目立体视觉系统中,深度信息的获得是分如下两步进行的。

(1) 在双目立体图像间建立点点对应。

(2) 根据对应点的视差计算出深度。

第一部分,也就是对应点问题,是双目立体视觉的关键;第二部分是摄像机模型问题。双目立体视觉模型中,双摄像机彼此参数一致,光轴平行且垂直于基线,构成一共极性(Epipolar)结构,这样做是为了缩小对应的搜索空间,只有水平方向的视差,简化了对应过

程，如图 5.18 所示。

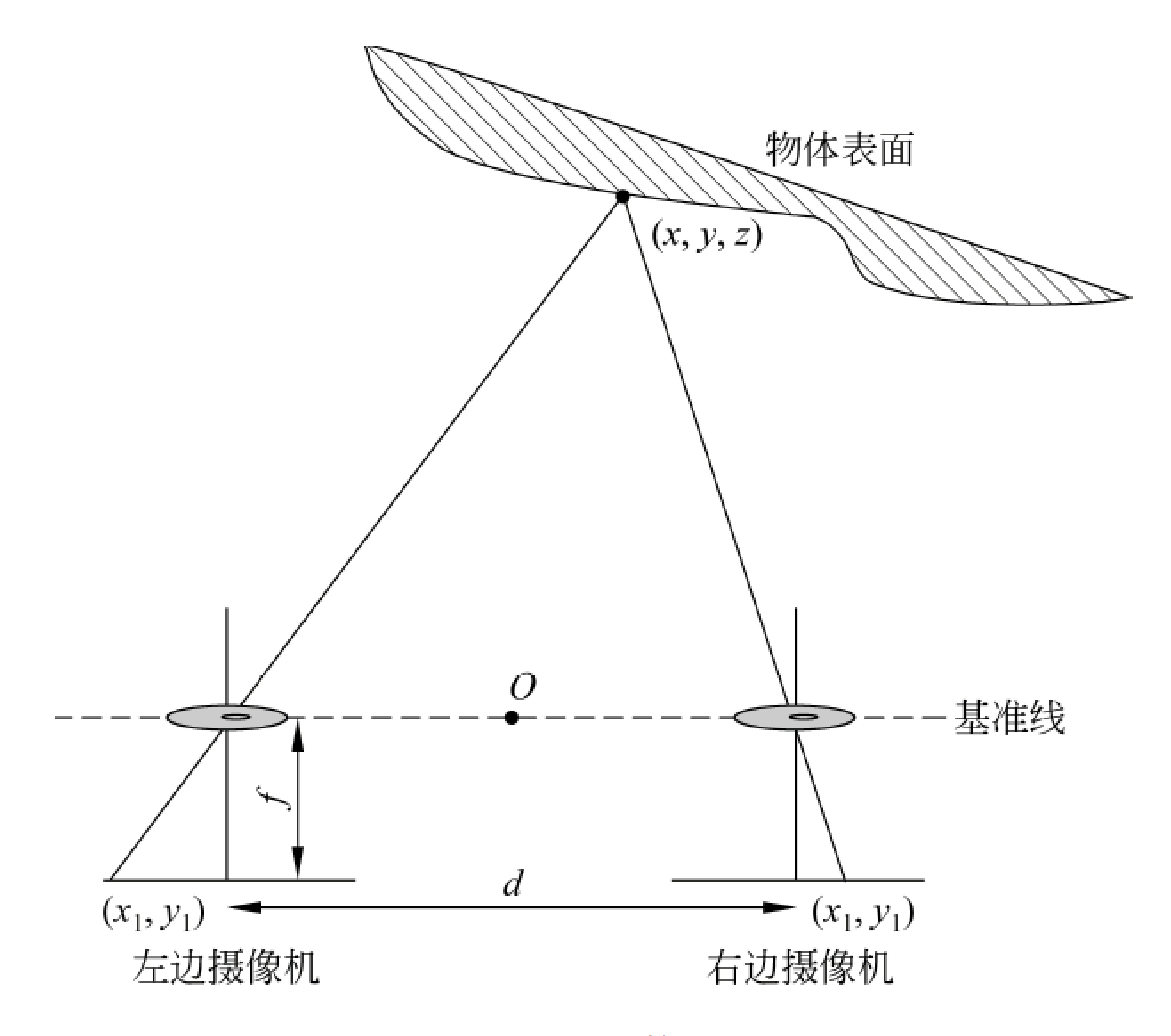



图 5.18　双目立体视觉系统

在图 5.19 中，设空间一点 $P(X,Y,Z)$ 在两个平行放置的完全相同的摄像机中像点分别是 $(x_1,y_1)$，$(x_2,y_2)$，则在知道基线长 $B$ 和焦距 $f$ 的情况下，可以计算出深度

$$Z = f\frac{B}{x_2 - x_1} = f\frac{B}{\Delta x} \tag{5.15}$$

这是双目立体视觉的基本原理，即根据视差来恢复立体信息。

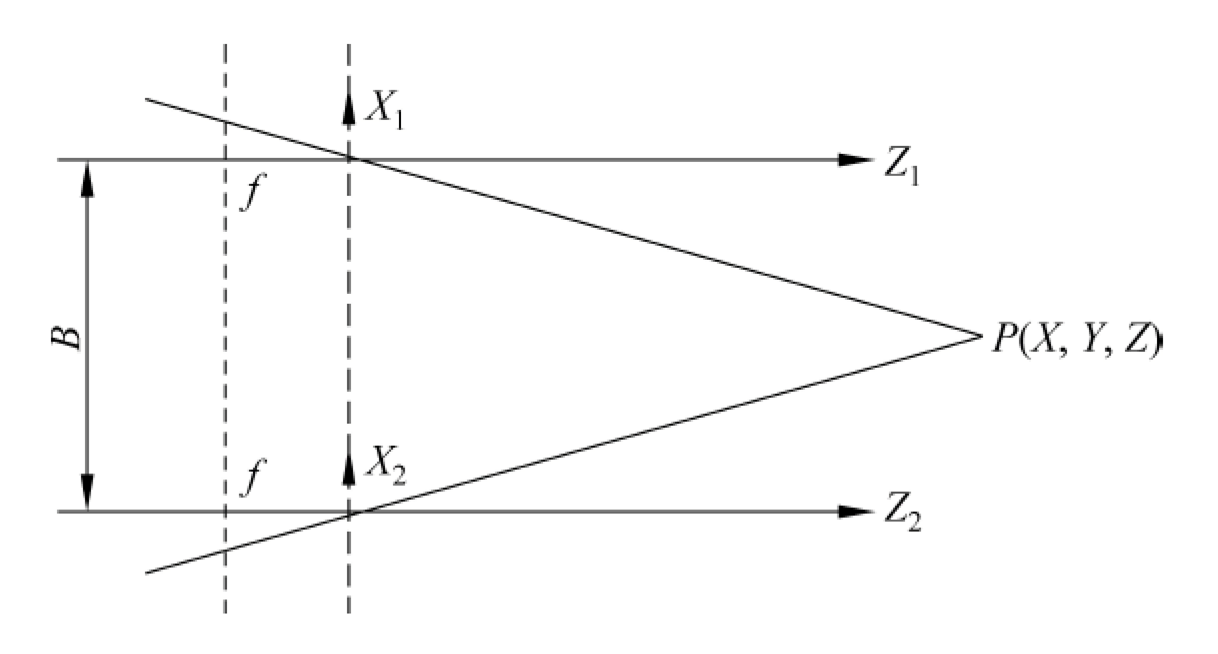



图 5.19　深度计算

## 5.8.5　利用启发式知识的方法

在图像理解中首先广泛利用启发式知识的是辜梓曼(A Guzman)[296]。他建立的 See 程序能把线画图分割成三维的物体。他的方法是对线画图中的由交汇的直线构成的顶点类型作分类。他发现根据顶点的类型就可以得到关于物体区域之间关系的局部线索。例如，Psi 类型的顶点(图 5.20)，指交汇在此顶点的几条线形成 ψ 形状)经常出现在两块柱状积木对准摆放时。这意味着上半部的两个区域属于一个物体，下半部的两个区域属于另一个物体。辜梓曼把属于同一物体的两个区域之间的连接称为链，并用链来表示关于区域之间连接关

系的启发式规则。See 程序根据这些关于顶点的启发式知识来连接各个区域。通过许多链连在一起的各个区域可能同属于一个物体,而属于不同物体的区域之间没有链或只有极少数的链连接。但这样的用于连接的启发式知识还不能最终确定哪个区域属于哪个物体。所以辜梓曼设计了根据区域之间链的数量、强度和拓扑关系来分配区域的启发式规则。他的算法可分析一幅相当复杂的线画图。这是早期视觉处理中应用启发式知识所取得的显著成就之一。它证明可以通过符号处理,而不是通过匹配过程来解释线画图。但是辜梓曼的方法还有根本性的困难。虽然 See 程序可识别三维物体,但所用的启发式知识仅限于二维的图像领域;其次,这些启发式知识是非常专门的,不通用并缺乏物理基础。这两个问题不是辜梓曼的方法所特有的,而是这类利用启发式知识方法的通病。

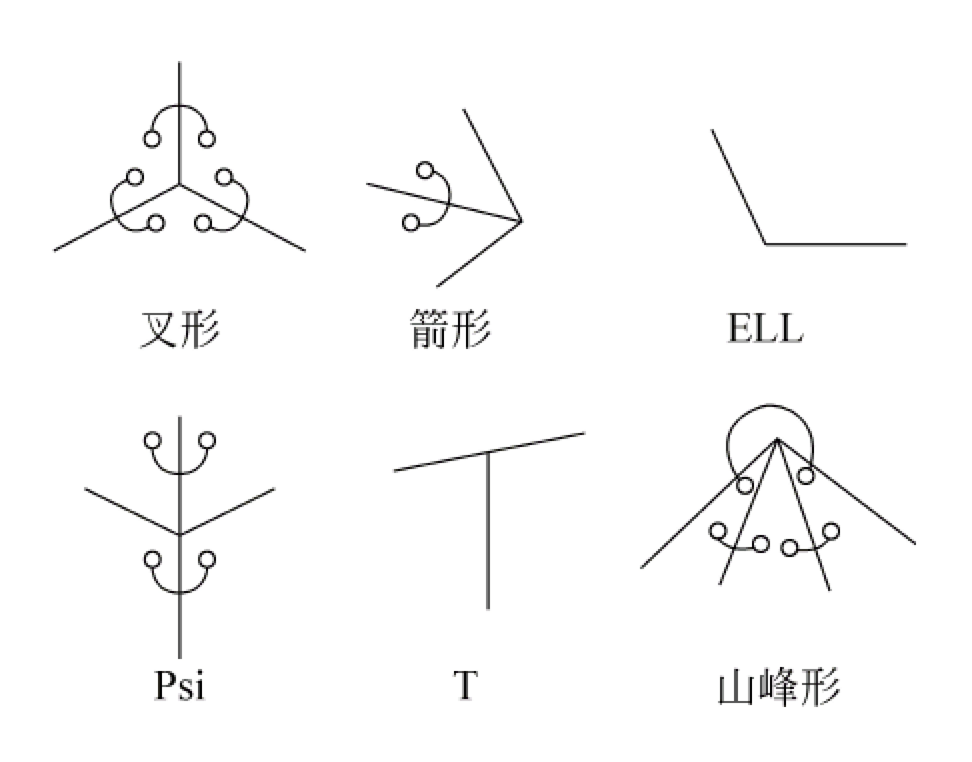



图 5.20 链围绕顶点的排列

# 5.9 同步化响应

## 5.9.1 概述

视觉信息处理是建立整体性概念的最恰当的范例,在视皮层的大部分区域,神经元处理的信息仅仅来自于视野的有限部分,也仅对特征构象的限定范围作出响应,因此,大量神经元的输出必须结合起来才能形成特定客体的完整表象。当存储的信息在记忆与回忆期间重新组合时也需要一种整体联结的机制。

在巴罗(H B Barlow)倡导的单细胞学说中[58],假定从初级处理阶段而来的输入集中到具有高度专一性响应特点的单细胞(即所谓“主教细胞”或“祖母细胞”),以解决联结问题,而客观世界的复杂多变性将使细胞数量出现“组合爆炸”式增长,无法获得实验的支持。在经过漫长的探索、实验和理性思维之后,研究工作重新返回到赫布的观点[311],即视觉客体由相互联结的神经元系综(Ensemble)来表象要比采用单细胞活性来表象更恰当。这种系综假定由对基本特性进行编码的神经元组成,复杂的表象在早期的处理阶段已形成,由于神经元在不同时间可以参与不同的系综,从而避免了巴罗学说面临的细胞数量爆炸式增长的困难。但是,根据 Hebb 的假设,某一时刻皮层的一个特定区域只有一个系综被激活,其他的系综被抑制,其结果是,如果神经元群 $A$ 表达特征 $a$,神经元群 $B$ 表达特征 $b$,当 $A$ 与 $B$ 重叠起来表达特征 $c$ 时,$a$ 与 $b$ 就消失了,这就是“叠加灾难”。此外,自然场景中的许多物体需要若干个系综共同激活且与视皮层的同一区域相对应,因而就无法确定哪些细胞属于哪个特定的系综,由此产生“捆绑问题”。

从信息处理的观点来看,无论是巴罗的假设还是赫布的观点,都是采用神经元脉冲平均发放速率编码外部刺激的。前者强调信息的串行处理,因而具有等级结构(即知识表达中联系上、下文的句法结构);后者强调信息的并行处理,自然就缺少等级结构,在信息的整合阶段,前者求助于“祖母细胞”,后者受困于“特征装配问题”。

究其根源，如何理解神经系统的编码方式就成为解决问题的关键和走出困境的途径[178]。将神经元仅仅看作是脉冲平均放电速率的积分器的概念，受到越来越多的新的实验事实的挑战，神经编码问题又重新唤起不同领域的科学家参与探索研究的热情。

马尔斯博格（Von der Malsburg）在整个 20 世纪 80 年代致力于神经编码问题的探索，他认为，系综应当用皮层神经元的同步发放来确定，而不应当用平均发放速率单纯的提高来确定，在这样的“相关模型”中，视象中知觉的相干性质通过相应的特征检测神经元的同步发放表现出来，与同一客体的特征相对应的神经元同步发放，而与不同客体相对应的神经元则以非相关的方式发放。本质上，这是时间相关假设[789]，依赖于动态联结方式来实现，神经元由轻微地改变其发放模式的瞬态关系而得以在系综之间迅速转换。换句话说，神经元群通过时间编码形成对客体的同步化响应[724]，处于同一系综之内的不同的神经元群用是否同步发放来标识。在嘈杂喧嚣的背景中，人们可以专心听到对方的谈话，这就是马尔斯博格引伸出的著名的“鸡尾酒会效应”（Cocktail-party Effect），是对相关理论的生动阐述[792]。

神经元群的同步振荡和同步化响应是继感受野与侧抑制概念之后重要的进展，通过理性的思维创立的这一新理论，正是迟迟才得到来自实验的支持而愈显其光辉，又一次证明了视-脑研究一刻也不能离开理论的指引。

1989 年时间相关理论出现了转机，格莱（C M Gray）等人[283]在猫的初级视皮层上发现了刺激诱发的神经元群活动的 $\gamma$ 同步振荡现象（频率为 30～70Hz），引起神经科学界的极大关注，其后相继进行的一系列神经生物学实验发现，这种振荡既存在于人脑、哺乳动物的脑、昆虫、蜗牛、蚯蚓的脑内，也存在于视觉系统（视网膜、外侧膝状体）和听觉系统之内；无论实验动物处于麻醉状态、清醒状态或是活动状态，情形均相同。由此可见，$\gamma$ 同步振荡是神经系统中的一种普遍现象。

视觉和脑研究从 1949 年赫布提出的成为经典的神经元群假设，转向 1972 年巴罗表述的“单细胞”观点，再到动态神经元集群的时空编码的假设，已走过了曲折漫长的探索之路，我们应当从中总结出必要而宝贵的经验教训，以真正科学的态度迎接神经网络、视觉和脑研究的新时期的到来。

### 5.9.2　神经生物学实验

$\gamma$ 振荡最初是在被麻醉的猫的初级视皮层上发现的，这种振荡在局部场电位中被观测到，通过微电极阵列记录加以验证，又在清醒的猫的视皮层做了同样的观测实验，用相干刺激激励时记录的神经元群具有零相位延迟的同步化，甚至空间相距达 7 mm 也是如此[283]。定位在不同视皮层区如 17、18 和 19 区，以及 17 区与后中侧超沟回（PMLS）之间的神经元群的同步化，即区域间同步化也在实验中观测到。如果不同皮层区的确代表了同一个客体的不同特征的话，特征的结合就可通过皮层区域间的同步化来实现。实验还证实，视觉对象的特征能在毫秒时间之内由特征检测神经元的同步化响应而彼此联系起来。响应同步化也应在大脑两半球出现，因为半球的视皮层直接接收来自视场半边的视网膜输入信号，为了使扩展到两个半视场的客体的特征结合起来，两半球之间的同步是需要的。实验同时记录大脑皮层左、右半球 17 区时发现了两半球响应同步化的出现，还发现了与同步化类似的强度与平均相位的关系[213]。

图5.21[214]是猫视皮层17区(纹状皮层)内相距7mm的两组神经元群在相干刺激(a),弱相干刺激(b)及非相干刺激(c)下的同步化振荡增强(d)、减弱(e)及消失(f)的实验结果。为了直观起见,神经元脉冲发放的测量数据用Gabor函数去拟合,并将在视皮层17区内微电极记录的结果(f)映射为二维的分布式神经元群(g),各自同步发放与同步振荡(图中分别用圆点和圆圈表示)。

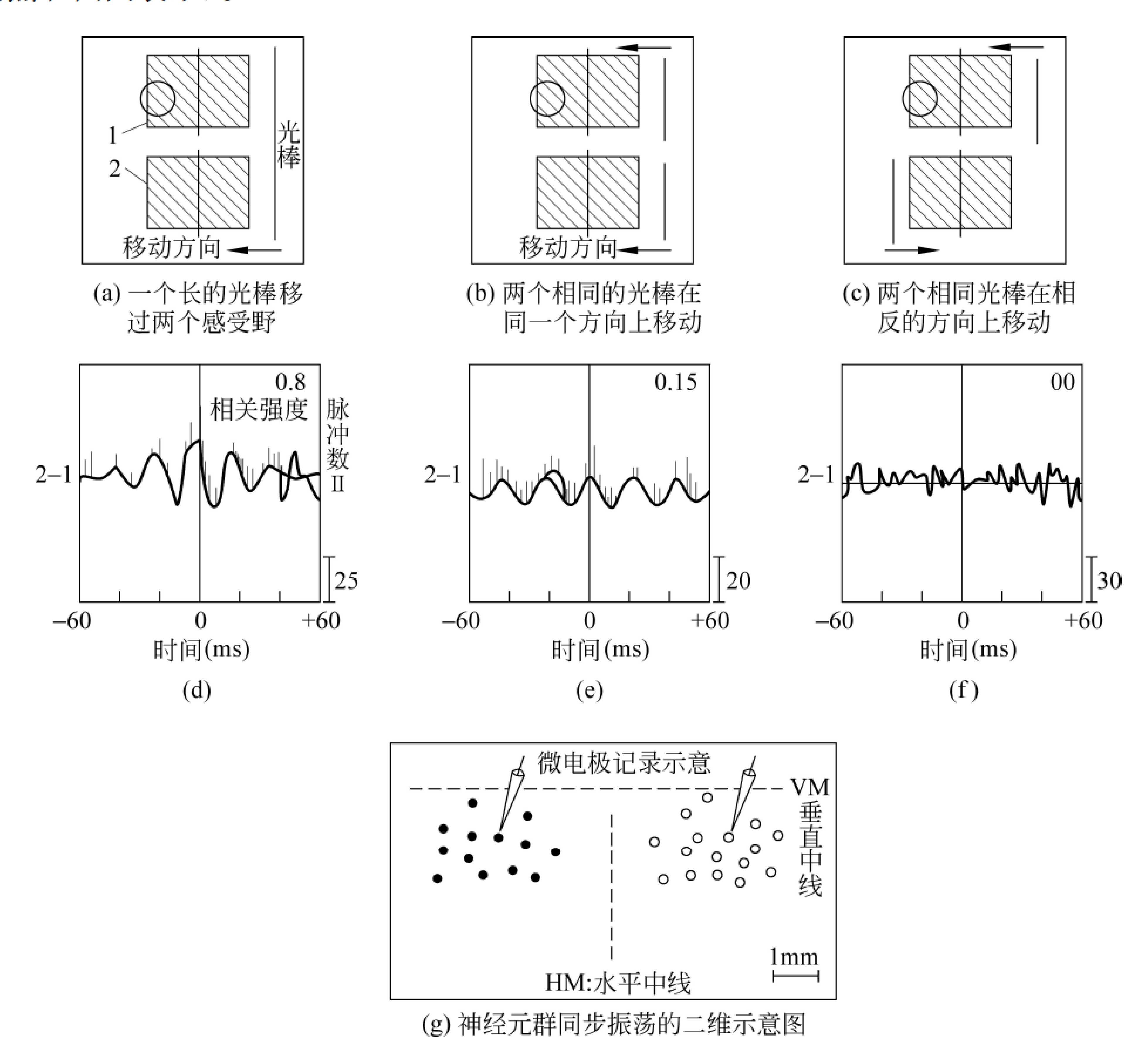



图5.21 猫视皮层17区内相距7mm两组神经元群对外界刺激(光棒)响应的实验结果

(注:图中的圆圈代表视场的中心)

### 5.9.3 时间编码

在神经元群同步振荡的与现象发现之后,研究工作集中到神经元是以什么方式编码外部世界客体的问题。图5.22[222]中这3种模式是表达了神经元编码中的同一个事件还是不同的时间模式表达了不同的事件?为了回答这个问题而进行的一系列实验表明,视网膜邻近的神经节细胞的同步发放确定了靠近它们的感受野之间重叠的区域,与单个神经节细胞的发放速率编码相比,视网膜可以用时间编码来建立视觉客体的更精确的表象,霍普菲尔德[332]提出的利用动作电位定时对模拟信息编码的模型支持神经网络处理相位编码信息比处理速率编码信息更加有效的观点。塞诺夫斯基指出[667],这种时间进程编码需要脉冲起始点的精确定时,也需要对来自不同神经元的脉冲的时间一致性高度敏感。视皮层

神经元的切片样品的实验测定表明，脉冲定时对同样的噪声电流的重复注入响应的精度少于 1ms，皮层回路能以很高的时间精度处理信息的实验证据不断增加，例如被克里克(F Crick)称做“Malsburg 突触”的是生理上的“快”突触，时间尺度在毫秒量级。里勒(A Riehle)等人建立的新的检测方法测定的神经元群的相干活性(或同步发放)的时间分辨率为毫秒量级[624]。

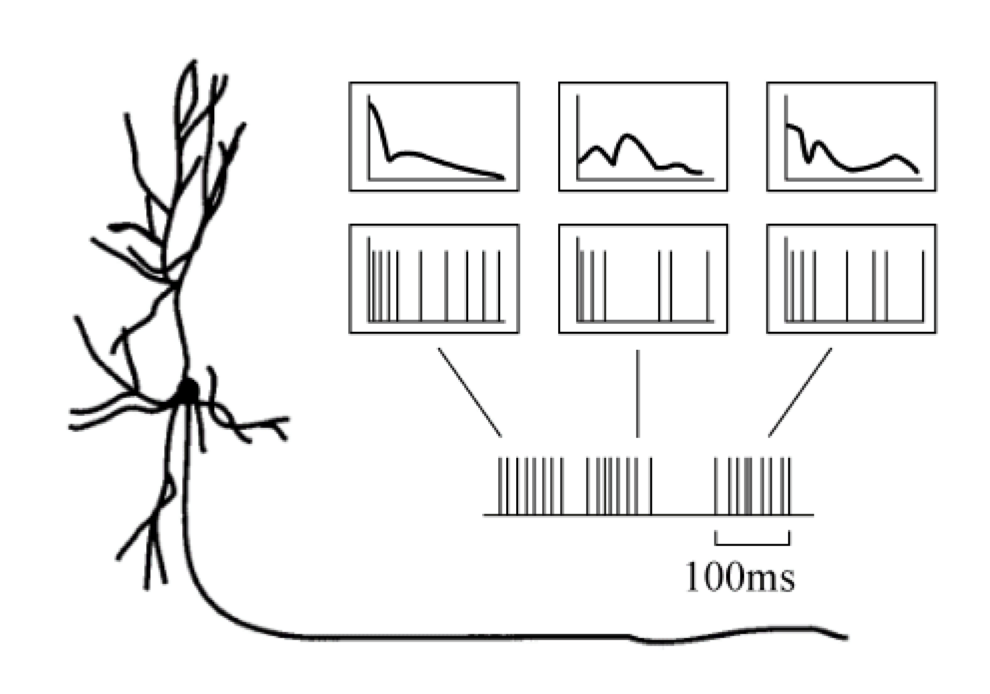



图 5.22　神经元轴突上记录的 3 组脉冲发放序列，平均发放速率相同(100Hz)而时间模式不同

在以往的研究中，几乎所有的神经网络模型都是从假设发放速率是输入信号求和后的 S 型函数开始的，大脑需要通过比较许多不同神经元的输出来确定刺激的精确性质；在时间编码中，一个神经元则是借助从众多时间输出模式的大清单中发出一个指令对刺激的单个特点的变化做出精确编码，γ 振荡则为时间编码提供了丰富的时间结构，而时间编码将会对并行的和分布式的信息处理的整合问题作出贡献。

## 5.9.4　视皮层的神经元振荡模型

为了深刻地理解 γ 振荡的生物学意义和在信息加工中所起的作用，几乎与神经生物学实验平行开展的是数学建模研究。

1990 年舒斯特(H G Schuster)等人建立了视皮层的神经元振荡模型[664]，其中传出突触既有兴奋性的也有抑制性的，在局域尺度上密集互联成模型神经网络，在更大范围内则是稀疏的，局部纤维束用兴奋性与抑制性神经元的平均活性来描述，相应的微分方程组确定了一个神经元振荡器，它能够由外部输入在活动与非活动态之间实现开关转换，研究两个这样的振荡器的耦合。舒斯特等人发现这种耦合特性是主动相关的，就是说，如果两个振荡器处于活动态，它们之间的耦合是紧耦合关系；如果一个或两个振荡器处于非活动态，则是弱耦合，与耦合关联的活动不同于直接相连的活动，直接相连是一种非动力学的固定连接。振荡器的特性是用极限环振荡器的相位而不是幅值进行描述的。

首先对视皮层内单个取向功能柱建立模型，做法是指定两个神经元群，按照它们的轴突末端是兴奋性的还是抑制性的加以区分，兴奋性神经元的激活率为 $e_k(t)$，抑制性的为 $i_l(t)$，总数分别为 $N_e$ 与 $N_i$，即 $k=1,2,\cdots,N_e$；$l=1,2,\cdots,N_i$。在脑结构上，功能柱内神经元彼此是互联的，这对应于振荡器之间的耦合，这样就可以对图 5.23 所示的功能柱建立相应的模型：

$$
\begin{cases}
\dfrac{\mathrm{d}e_k}{\mathrm{d}t} = -e_k + S\left[a_e\left(\dfrac{1}{N_e}\sum_{i=1}^{N_e}U_{kl}e_l - \dfrac{1}{N_i}\sum_{i=1}^{N_i}V_{kl}i_l - \theta_k^e + p_k\right)\right], k = 1,2,\cdots,N_e \\
\dfrac{\mathrm{d}i_k}{\mathrm{d}t} = -i_k + S\left[a_i\left(\dfrac{1}{N_e}\sum_{i=1}^{N_i}W_{kl}e_l - \dfrac{1}{N_i}\sum_{i=1}^{N_i}Z_{kl}i_l - \theta_k^i\right)\right], k = 1,2,\cdots,N_i
\end{cases}
\tag{5.16}
$$

为了减少变量数目,上述方程用平均场方程代替

$$
\begin{cases}
\dot{E} = -E + S[a_e(c_1E - c_2I - \Theta^e + P)] \\
\dot{I} = -I + S[a_i(c_3E - c_4I - \Theta^i)]
\end{cases}
$$

$$
\begin{cases}
E(t) = \dfrac{1}{N_e}\sum_{i=1}^{N_e}e_i(t), I(t) = \dfrac{1}{N_i}\sum_{i=1}^{N_i}i_l(t), \Theta^e = \dfrac{1}{N_e}\sum_{k=1}^{N_e}\theta_k^e, \Theta^i = \dfrac{1}{N_i}\sum_{k=1}^{N_i}\theta_k^i \\
P = \dfrac{1}{N_e}\sum_{k=1}^{N_e}P_k, S(x) = \dfrac{1}{1+e^{-x}}
\end{cases}
\tag{5.17}
$$

这种代替直到 $1/\sqrt{N_i}$ 阶都是正确的($N_i \approx 10^4$),对 $S(x)$ 作泰勒展开与变换就可得出耦合振荡的相位方程组:

$$
\begin{cases}
\dot{\varphi}_1 = \omega_1 - k_{12}\sin(\varphi_1 - \varphi_2) \\
\dot{\varphi}_2 = \omega_2 - k_{21}\sin(\varphi_2 - \varphi_1)
\end{cases}
\tag{5.18}
$$

式(5.17)中 $I$ 表示抑制性神经元,$E$ 表示兴奋性神经元,耦合用 $c_1, c_4$ 表示,符号+与-分别表示兴奋性与抑制性耦合,$P$ 是外部输入[664]。

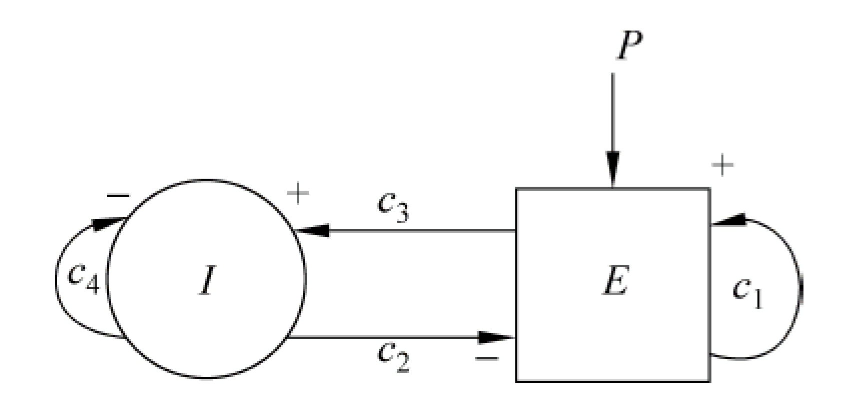



图 5.23 两个耦合的神经元群组成的功能柱示意图

显然,相位方程的状态变量数目大为减少,使数值积分容易进行,数值实验的结果成功地模拟了格莱等人在猫的视皮层上发现的同步振荡现象。

舒斯特等人建立的模型为什么能获得成功?主要原因是这个模型不仅是以神经生物学的事实为基础,如神经元群分成兴奋型与抑制型;而且也以视皮层的典型构造为基础,就功能柱而言,模型神经网络是非局域的,对于超级功能柱这种大尺度耦合的网络,模型则是局域的和稀疏的,对应于相距达 7mm 的皮层神经元群的长程同步振荡的情形,模型以神经元群之间耦合的非线性振荡器的相位描述为基础,因而能够探讨同步振荡与脉冲定时编码视觉客体的关系。

这个模型的深刻意义还在于说明了数学建模与仿真计算对于视觉信息处理研究的重要性。

## 5.9.5 视觉系统中的表象与尺度变换

马尔在创立视觉计算理论时,一再强调表象概念的极端重要性,因为把视觉看作是一个完善的信息处理系统的观点已广为人们接受[462]。但是,视觉不是一般意义上的信息处理系统,

表象是视觉系统作为智能信息处理的独有特点，是视觉计算理论的核心概念，它规定了视觉信息处理所要解决的是什么问题以及能使该问题获解的、由外部世界的一般性质所决定的约束条件，从这个意义上说，所谓视觉，就是在计算理论限定的范围内，对各种视觉信息处理模块选取合适的表象和算法，这里所指的计算，就本质而言是一种模拟，并通过相应的硬件（脑或计算机）来实现，从而认知外部世界中有什么东西，它们在空间所处的位置如何。

马尔提出的三级内部表象是：要素图（图像的表象），2.5 维图（可见表面的表象）和三维模型表象（用于识别三维物体形状的表象），其中最关键的是 2.5 维图。在一般情况下，物体形状的分层次结构的描述与识别是和人的知识、经验、学习与记忆有关的非常复杂的信息处理过程，自然会受到脑的其他部位的影响。

选择图形的局部几何性质，如线段、平行性、相对位置和朝向等作为视觉要素图的基本信息，是由于哺乳动物大脑皮层存在简单细胞、复杂细胞所构成的感受线条和边界的特征检测器以及大脑皮层功能柱能对来自视网膜、外侧膝状体的输入进行重新组合，分析刺激的线条、轮廓、方向、运动等空间特性；也是由于计算机具有适于计算局部几何性质的功能。

马尔明确地指出，视觉的一个处理过程是从一种表象获得另一种表象的一个映射；对于人类视觉这一特定的情形，原始表象无疑是由视网膜上的光感受器所检测到的图像强度值阵列所组成的。

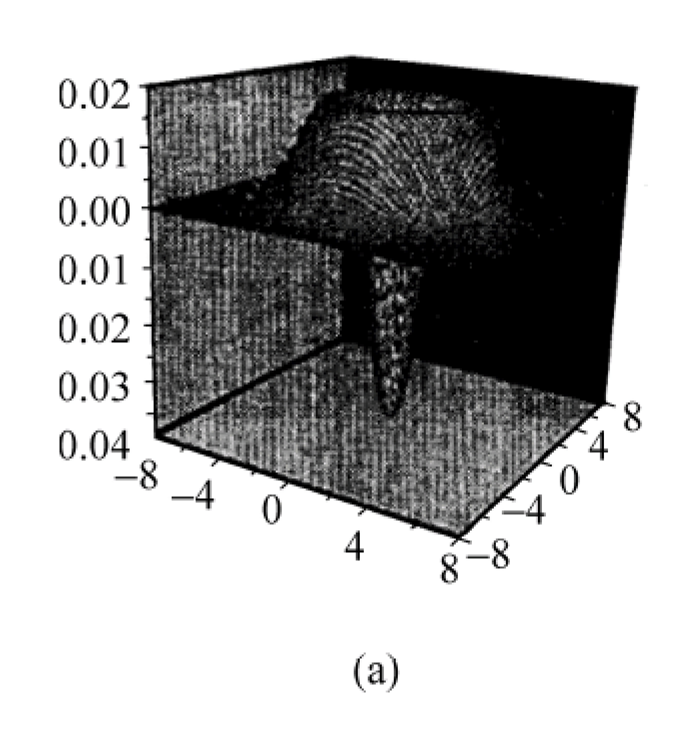
这一理论经受住了创建之后一系列实验与研究进展的检验，成为探讨在视觉系统中真实的信息处理过程的严密的理论，它受到的诘难主要是起因于 2.5 维图，尽管波杰奥（T Poggio）发展了一整套关于初级视觉的正则化方法，通过选取一个能使某一恰当的泛函极小化的函数来限定可以接受的解空间，消除在哈达玛德（Hadamard）意义下从图像的光强分布恢复物体三维形状的不适定问题，这对于计算机视觉无疑是重要的，但对于理解生物视觉而言，就未必如此。面对这种越来越复杂的数学化的抽象处理，人们不禁要问，这是视觉中的真实过程吗？

大脑皮层上千亿神经元彼此以特殊、复杂而准确的方式联结成网络，其目的是在感知外部世界时神经元的信息处理与计算方式能以“多”取胜，而不是以“复杂”取胜，视觉系统凭靠数目众多的神经网络处理信息时是简单、和谐和美的统一，复杂化的结构体现了简单化的准则，适应了变化万千的客观世界。

同步振荡现象的发现和神经元脉冲发放具有毫秒量级的分辨率，为重新研究和检验正则化理论以及为同样的目的提出的其他相关理论创造了条件。

视觉的生物组织，外部客体的特征以及成象过程均具有层次结构，因此视觉表象必须能以多种尺度把图像的性质表达清楚。马尔虽然给出了由高斯函数的二阶微分组成的算子$^2G(r,\sigma)=-(1/\pi\sigma^4)(1-r^2/2\sigma^2)\exp(-r^2/2\sigma^2)$，其形状如墨西哥草帽，作为视觉通道的基本运算，但这个算子的积分值不为零，不能真正成为模拟 on 型和 off 型感受野特性的算子，方差 $\sigma$ 所起的作用类似于尺度因子。在小波变换（Wavelet Transform）出现之后，就可以用多尺度或多分辨分析（Multiscale 或 Multiresolution Analysis，MSA 或 MRA）将视觉表象的尺度作统一的数学描述，引入尺度因子 $a$，利用高斯函数就可以构造出马尔子波，它的积分值为零，因此是一个真正的子波，两者的比较如图 5.24 所示。

$$\frac{1}{a^2}\psi\left(\frac{r}{a}\right)=\nabla^2 G\left(\frac{r}{a}\right)=-\frac{1}{2\pi a^2}(1-r^2/a^2)\exp(-r^2/2a^2) \tag{5.19}$$



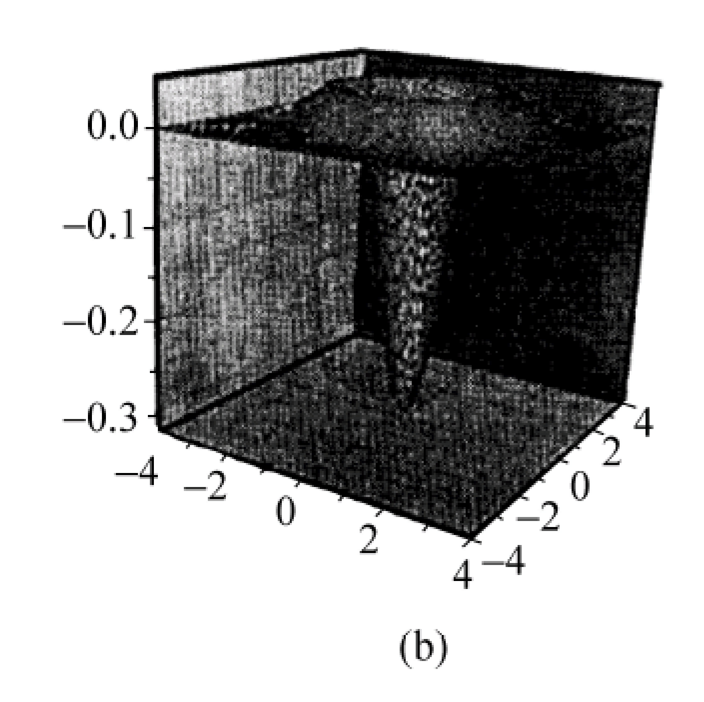




图 5.24　$\nabla^2 G\left(\frac{r}{a}\right)(a)$与$\nabla^2 G(r,\sigma)$的二维形状

显然 $\sigma$ 或 $a$ 所起的作用在频域相当于将视觉通路划分为不同频带的空间频率通道,这已为开普贝尔(Campbell)和威尔森(Wilson)等人经典的研究与实验结果所证实[462],用小波变换的分解与综合能更深刻地反映初级视觉信息处理过程,如图 5.25(a)所示;图 5.25(b)是图像在时域中的多尺度分析过程[509]。

当我们把图 5.25 的信息处理过程表示成子波变换时,则有

$$\int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)\mathrm{d}x=\int_{-\infty}^{\infty} f(ax)\sqrt{a}\psi\left(x-\frac{b}{a}\right)\mathrm{d}x, a>0, f\in L^2(R) \quad (5.20)$$

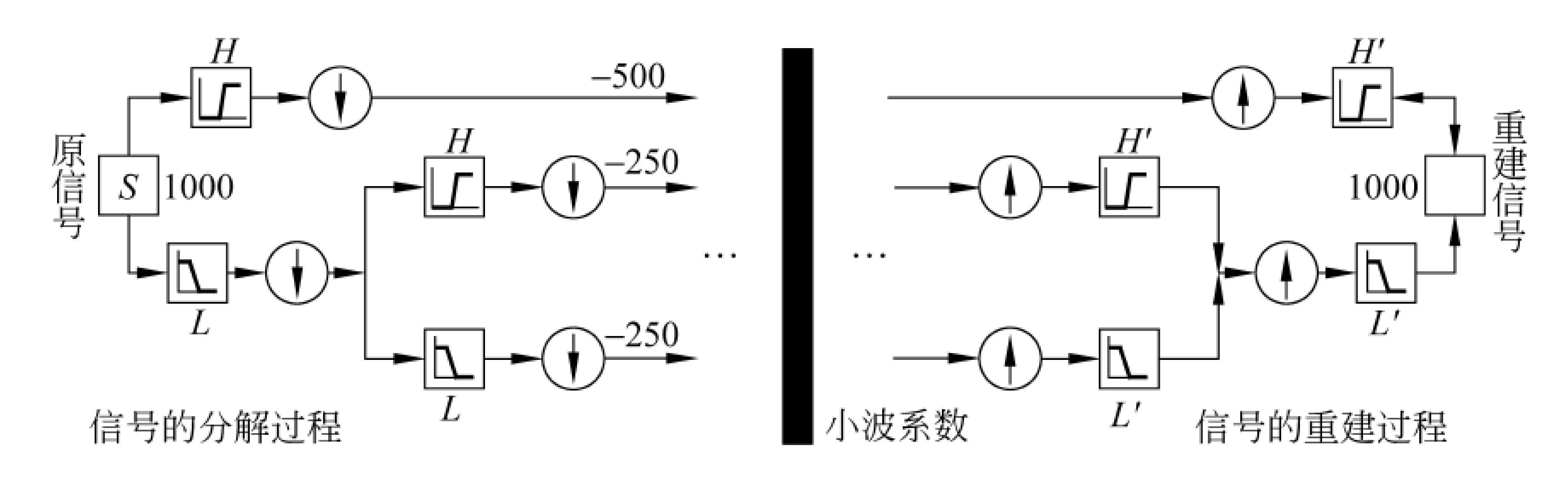
其实,式(5.20)表示了视觉的这样一个基本事实,即当尺度 $a$ 增大时,$\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)f(x)$表示减小视锐度($1/\sqrt{a}$)与增大对比度[$\psi(x/a)$]以观察分析一个支撑区为 $D$ 的对象 $f(x)$;而 $\sqrt{a}\psi(x-b/a)f(ax)$则表示增大视锐度($\sqrt{a}$)与减小对比度[$\psi(x)$]以观察支撑区缩小了 $a$ 倍(即 $D/a$)的同一个对象 $f(ax)$,两者的效果是一样的,它们均使变换前后的图像像素密度保持不变,体现了人类视觉感知自然景物的恒常性与多尺度分辨过程。

因此,用子波变换这一有效的数学工具对视觉系统的表象作深入的多尺度分析,有助于视觉计算理论的发展与完善,是当前受到重视的一个研究方向。

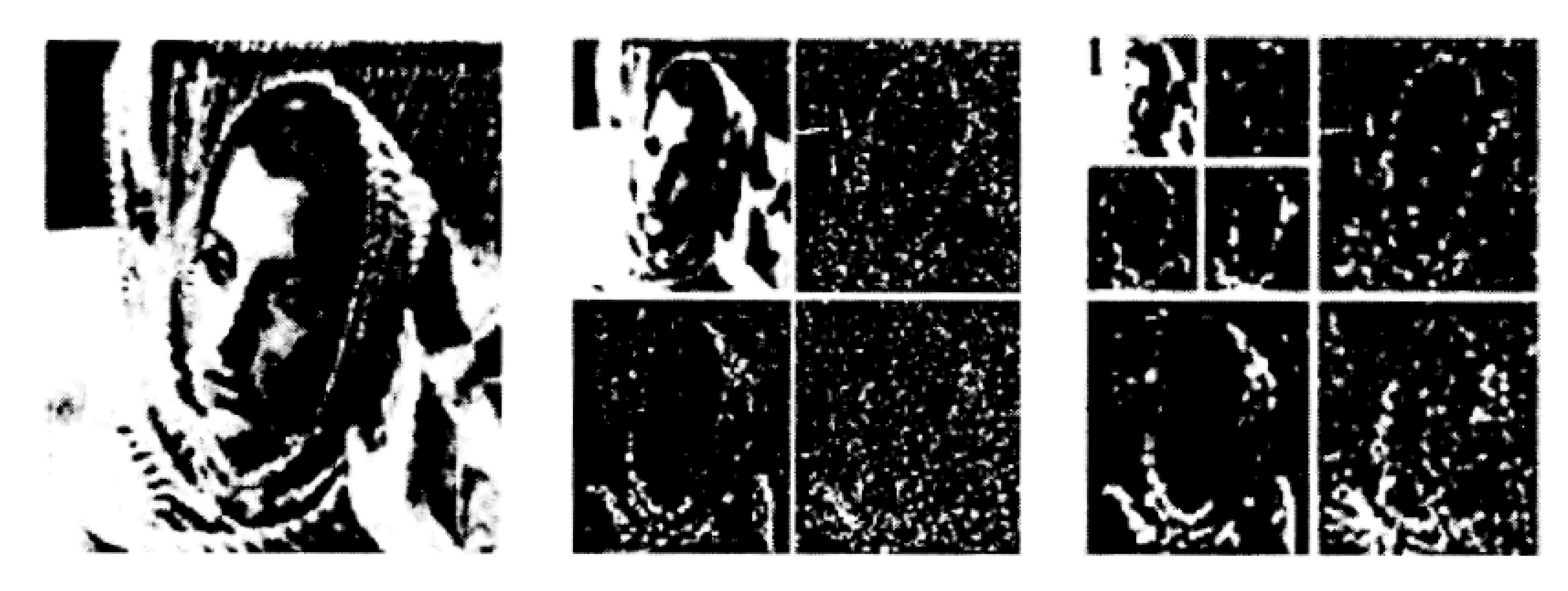
### 5.9.6　神经网络中的非线性动力学问题

在视-脑研究中神经网络的观点是一种全局观点,无论就单个神经元的 S 型传输特性而言,还是就神经元作为振子而言,用非线性动力学方法进行分析都可以在网络的全局行为或者演化的终态方面获得深刻的认识。20 世纪 80 年代初期,霍普菲尔德的神经网络集合运算模型是为解决神经网络如何实现最优化算法并通过硬件实现的课题而提出的,其最大特点是采用能量函数极小化作为神经网络的最优运算结果,从而与联想记忆问题联系起来。能量是一个普适量,也是系统整体行为的度量,维持最小能耗是生命系统进化的结果,因而用能量的泛函取极小值描述大脑皮层的整体运算就更接近于脑内发生的真实过程。

1995 年霍普菲尔德在神经元群同步化响应的基础上提出了用动作电位定时编码来表象外部刺激的模式识别原理,他建议用时间延迟网络和径向基函数计算其表象。这些研究推动着神经网络研究的进展。神经网络是典型的非线性系统,就是说,视皮层神经元是“成




(a) 多尺度的频域滤波过程，S为信号；L、L′为低通滤波；H、H′为高通滤波；↓为欠采样，↑为过采样



(b) 时域中图像的多尺度分析

图 5.25　视觉系统中表象的多尺度表示

注：尺度 $a$ 按二进方式增加

簇”发放的，它们通过复杂的联结形成的神经网络的行为是高度非线性的。无论是视觉环境，还是脑内环境都存在大量背景噪声，通过随机共振对输入信号作出优化响应，从而赋予噪声在神经网络系统中的功能作用，由 $N$ 个神经元组成的求和网络对于神经生理的感知系统刺激的响应如图 5.26 所示，它对视觉系统检测信号时噪声的作用的研究具有重要的参考价值[332]。

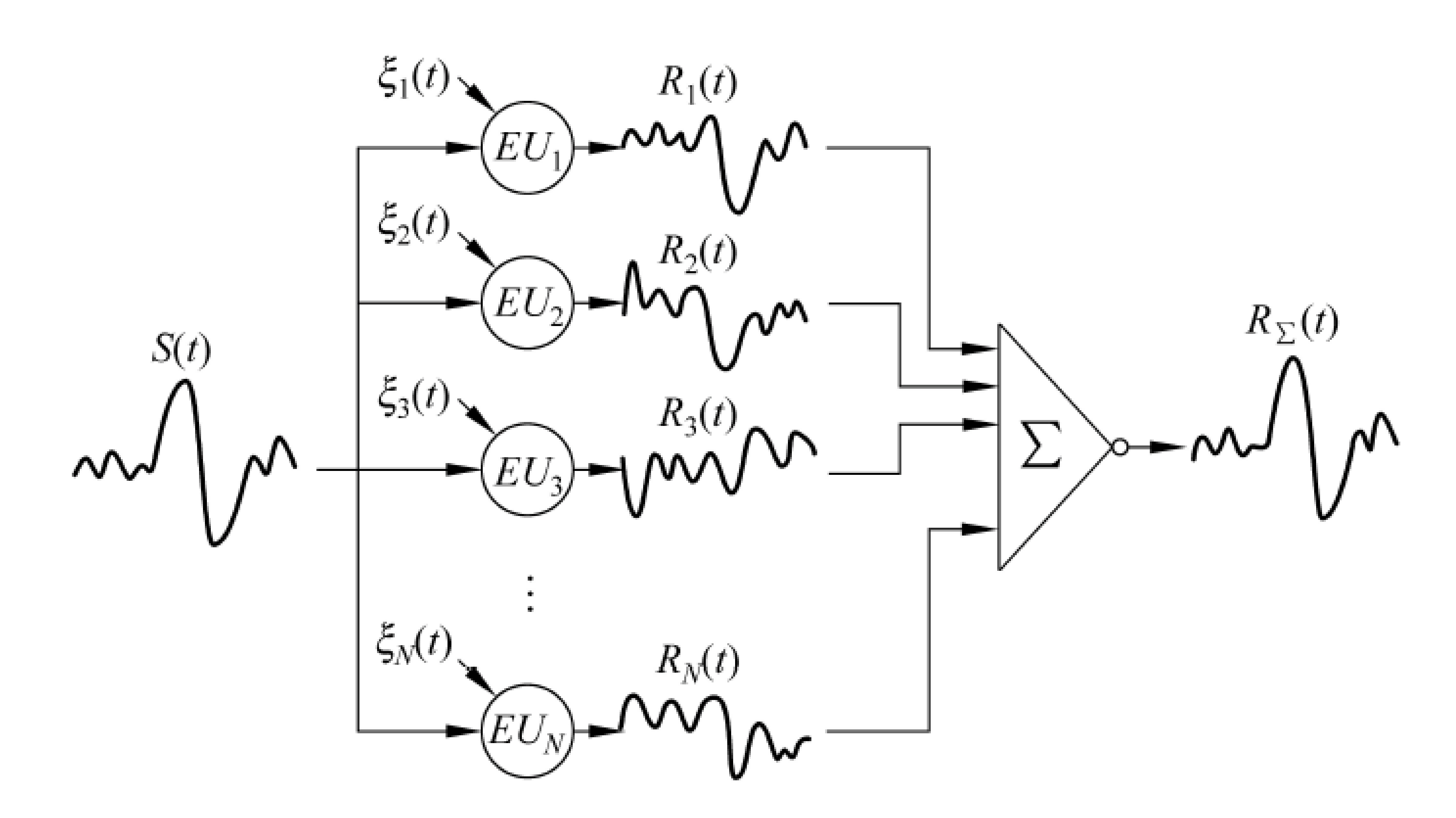



图 5.26　由 $N$ 个神经元组成的求和网络对输入信号的响应

由 $N$ 个 Fita Hugh-Nagumo 模型神经元 $EU_i$ 组成的求和网络对输入信号的响应神经元遵从下列方程：

$$
\begin{cases}
\varepsilon \dot{V}_i = V_i(V_i - a)(1 - V_i) - W_i + A + S(t) + \xi_i(t) \\
\dot{W} = \dfrac{\mathrm{d}w}{\mathrm{d}t} = V_i - W_i - b
\end{cases}
\tag{5.21}
$$

式(5.21)中,$\varepsilon=0.005$;$a=0.5$;$b=0.15$;$A$ 是恒定的调谐信号;$S(t)$是慢变化的非周期信号,$V_i(t)$是快变量,$W_i(t)$是慢变量,$R_i(t)$是每个单元的平均发放速率信号,$\xi_i(t)$是均值为零的高斯白噪声。

当前,神经元群同步振荡现象的发现深刻地改变了人们对网络的观点,对于动态的或功能性联结的神经网络,同步振荡是一种选择机制,它根据外部刺激选择出适合的动态神经元集群进行表象处理,当外部客体变化了,瞬态形成的神经元群就由新形成的集群来代替,里切蒙德(B J Richmond)等人采用 Walsh 功能图研究过视觉信息处理中的“旋转不变性”识别问题[623],科能(K W Konen)等人则采用动态联结网络来解同一问题[385]。

# 第 6 章<<

# 听觉感知

声波作用于听觉器官,使其感受细胞兴奋并引起听神经的冲动发放传入信息,经各级听觉中枢分析后会产生听觉。听觉过程包括机械→电→化学→神经冲动→中枢信息加工等环节。从外耳的集声至内耳基底膜的运动是机械运动,毛细胞受刺激后引起电变化,化学介质的释放、神经冲动的产生等活动,冲动传至中枢后则是一连串复杂的信息处理过程。

## 6.1 听觉通路

言语听觉比我们想像的要复杂得多,部分原因是口语速率最高达每秒 12 个音素(基本口语单位)。我们能理解的口语速度最多不能超过每分钟 50～60 个语音。在正常口语中,音素会出现重叠现象,同时存在一种协同发音现象,即一个语音片断的产生会影响到后一个片断的产生,而线性问题是指协同发音引起言语知觉困难的现象。与线性问题相关的问题是非恒定性问题。这一问题是因任何给定的语音成分(如音素)的声音模式并不是恒定不变的而引起的,而是它受到前后一个或多个声音的影响。这对辅音来说更是如此,因为它们的声音模式常常依赖于紧随其后的元音而定。

口语一般由连续变化的声音模式以及少数停顿所组成。这与有独立声音构成的言语知觉形成鲜明对比。言语信号的连续性特征会产生分割问题,即决定一个连续的声音流怎样被分割成词汇。

从耳蜗到听觉皮质的听觉系统是所有感觉系统通路中最复杂的一种。听觉系统的每个水平上发生的信息过程和每一水平的活动都影响较高水平和较低水平的活动。在听觉通路中,从脑的一边到另一边有广泛的交叉(见图 6.1)。

进入耳蜗神经核后,第八对脑神经听觉分支纤维终止于耳蜗核的背侧和腹侧。从两个耳蜗核分别发出纤维系统,从背侧耳蜗发出的纤维越过中线,然后经外侧丘系上升到皮质。外侧丘系最后终止于中脑的下丘,从腹侧耳蜗核发出的纤维,首先与同侧和对侧的上橄榄体复合体以突触联系,上橄榄体是听觉通路中的第一站,在这里发生两耳的相互作用。

上橄榄复合体是听觉系统中令人感兴趣的中心,它由几个核组成,其中最大的是内侧上橄榄体和外侧上橄榄体。根据几种哺乳动物的比较研究,发现这两种核的大小与动物的感觉能力之间的相互关连,Harrison 和 Irving 指出这两种核有不同的机能。他们指出,内侧上橄榄体和关联到眼球运动的声音定位有关,凡具有高度发展的视觉系统以及能注视声音的方向而作出反应的动物,内侧上橄榄核有着显著的外形。而另一方面,他们推论外侧上橄榄体则与独立于视觉系统以外的声音定位有关。具有敏锐的听觉但视觉能力有限的动物,都有显著的外侧上橄榄核。蝙蝠和海豚的视觉能力有限,但有极其发达的听觉系统,完全没有内侧上橄榄核。

从上橄榄复合体出发的纤维上升经过外侧丘系到达下丘。从下丘系将冲动传达到丘脑

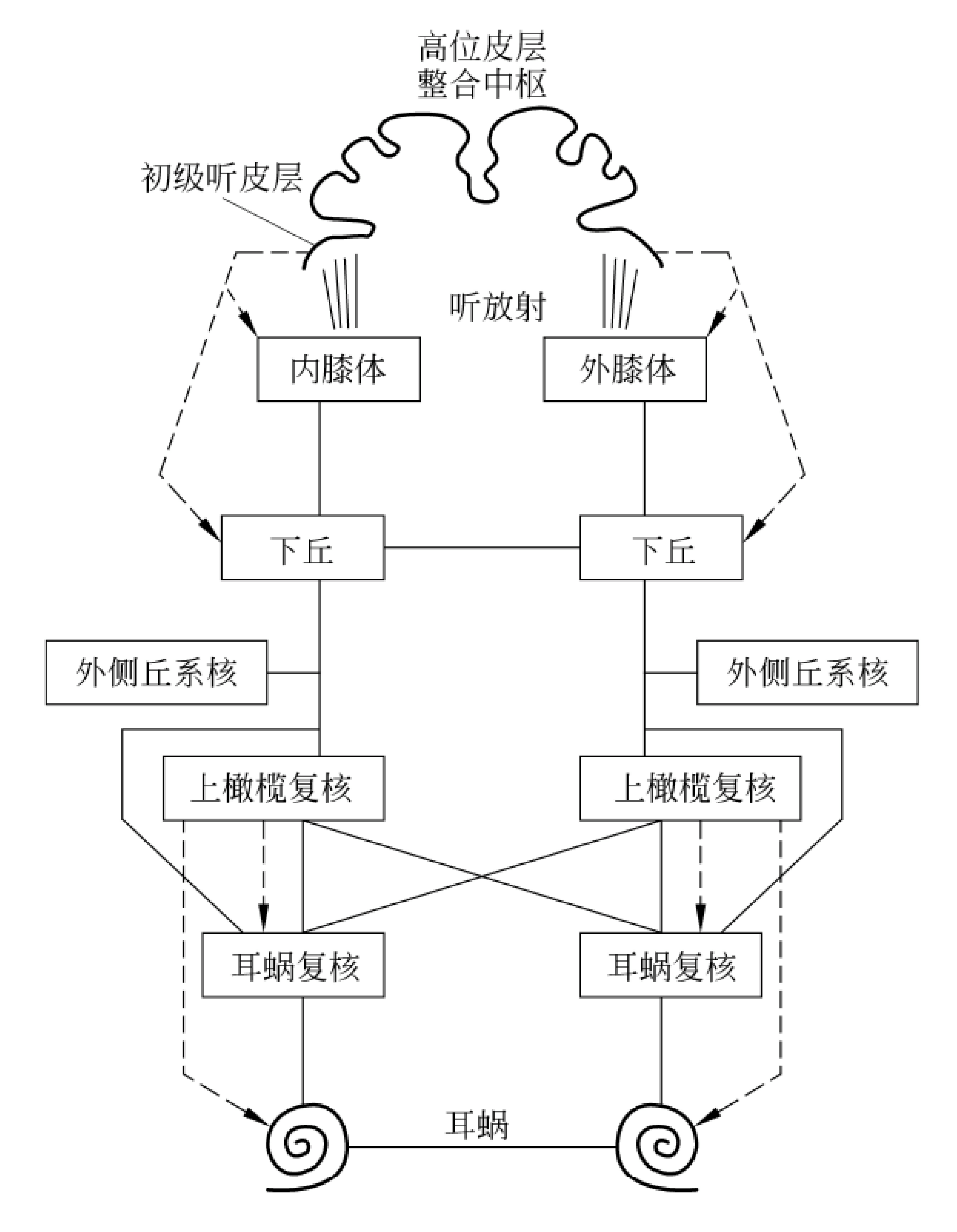



图 6.1 听觉通路

的内侧膝状体。连接这两个区域的纤维束,叫做下丘臂。从内侧膝状体,听觉反射的纤维将冲动传导颞上回(41 区和 42 区),即听觉皮质区。

1988 年伊里斯(A W Ellis)和杨(A W Young)提出了一个口语单词加工的模型(参见图 6.2)[211]。这个模型包括如下 5 个成分。

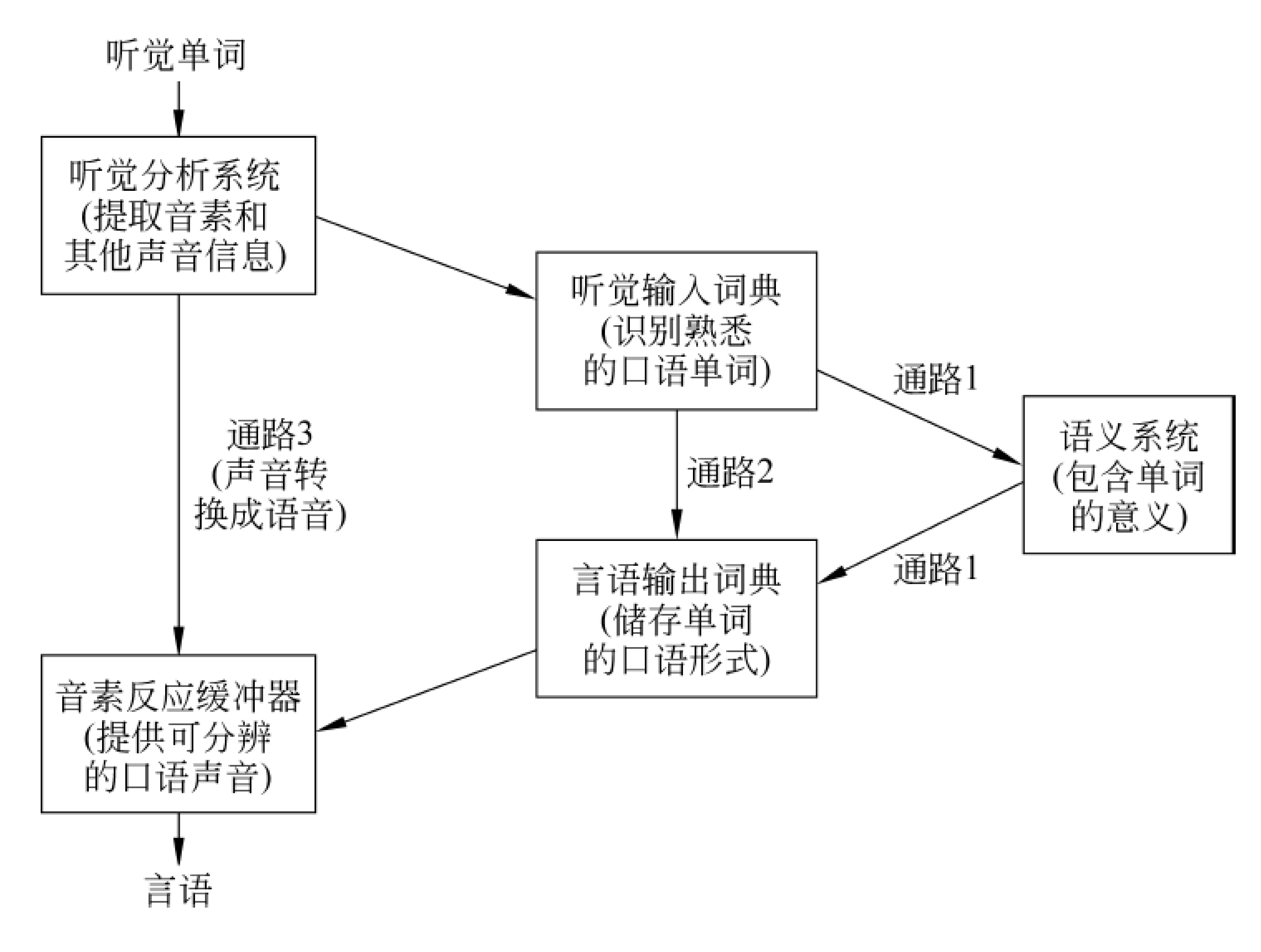



图 6.2 口语通路模型

（1）听觉分析系统：用于从声波中提取音素和其他声音信息。

（2）听觉输入词典：包含听者知道的关于口语单词的信息，但不包含语义信息。这个词典的目的就是通过恰当地激活词汇单元来识别熟悉单词。

（3）语义系统：词义被储存于语义系统之中。

（4）言语输出词典：用于提供单词的口语形式。

（5）音素反应缓冲器：负责提供可分辨的口语声音。

这些成分可以各种方式组合起来，因此在听到一个单词至说出它之间存在三条不同的通路。

（1）通路 1

这条通路利用听觉输入词典、语义系统和言语输出词典。它代表了无脑损伤人群正常识别和理解熟悉单词的认知通路。如果一个脑损伤患者只能利用这条通路（也许加上通路 2）的话，那么他将能够正确地说出熟悉单词。然而，在说出不熟悉单词或非词时将出现严重困难，因为这类材料没有存储于听觉输入词典之中。在这种情况下，患者需要使用通路 3。

（2）通路 2

如果患者能够使用通路 2，但通路 1 和 3 受到严重损伤，那么他们应该能够重复熟悉单词，但不能理解这些单词的意义。此外，患者也应该存在对非词的认知障碍，因为通路 2 不能处理非词信息。最后，由于这些患者将使用输入词典，所以他们应该能够区分词与非词。

（3）通路 3

如果一个患者只损伤通路 3，那么他或她将展示在知觉和理解口语熟悉单词方面的完好的能力，但在知觉和重复不熟悉单词和非词时会出现障碍。这种情况临床上称之为听觉性语音失认。然而，他阅读非词语时的能力完好。

## 6.2　听觉信息的中枢处理

中枢的信息处理过程甚为复杂，目前对它还缺乏较全面的理解。中枢神经系统的一般神经生理活动规律，有不少应是在听觉中枢活动中同样起作用的。从耳蜗神经传入的冲动，在时间和空间上同所接受的声音特性不同而有不同的构型，这是输入信息编码的总形式。最后产生的听觉，一般能准确而精细地反映声音的各种复杂特性，这也是清楚的。夹在输入输出之间的中枢信息处理过程这一个黑匣子，正是当前听觉生理研究的核心问题。

### 6.2.1　频率分析机理

自从一百多年前，亥姆霍兹提出了共振学说以来，不同的年代不同的作者提出过多种学说，解释耳蜗的频率分析机理。意见分歧很大，但基本上可用两种观点进行概括：一种观点认为不同频率的声音兴奋基底膜不同部位的感受细胞，兴奋部位是频率分析的依据；另一种观点认为不同频率的声音使听神经兴奋后发放不同频率的冲动，冲动频率是声音频率分析的依据。前一种观点称为部位原则或部位学说，后一种观点称为时间原则或冲动频率学说。两种观点都有事实依据，但也各有不足之处。现在人们普遍认为它们二者不是互相排斥，而是可以互相补充的。听觉的频率分析不是一个简单的周边过程，中枢在精确的声音辨

别中起决定性的作用。

### 1. 行波论

贝克西(Bekesy)在显微镜下对动物耳蜗的直接观察,及在他所设计的耳蜗模型上的研究都表明,声音引起基底膜波动的行波确是从耳蜗基部开始逐步向蜗顶移动的,在移动过程中行波的振幅是变化的,振幅最大点的位置及行波移动的距离都随声音的频率变化,振幅最大点在高频刺激时靠近耳蜗基部,频率逐渐降低时它便逐渐离开基部朝蜗顶方向移动,低频时靠近蜗顶。经过了最大点后振幅便很快衰减。行波振幅最大处螺旋器感受细胞受到的刺激最强。按照部位原则,行波论认为行波振幅最大点位置是对声音频率分析的依据,基底膜靠基部处接受高频声刺激,靠蜗顶处接受低频声刺激,当中按频率高低次序排列。因此基底膜就成为一个初级的频率分析器。

还可以用微电极分别记录不同的听纤维的频率调谐曲线,每条曲线表示单条听纤维的阈值与频率的关系,并且都有一个最小阈值,此最小阈值所对应的频率都是该束纤维的特征频率。

### 2. 排放论

单根神经纤维重复发放冲动的能力有限,一般在每秒数百次以下,跟不上较高的声音频率,这是早年冲动频率论受到非议的主要论据。韦佛(Wever)提出排放论,对此作出较满意的解释。排放论的解释是:若多根纤维随声波的周期而同步地轮流发放,则每一根纤维发放的频率不要很高,总体纤维上冲动排放却可跟上很高的频率(图6.3)。

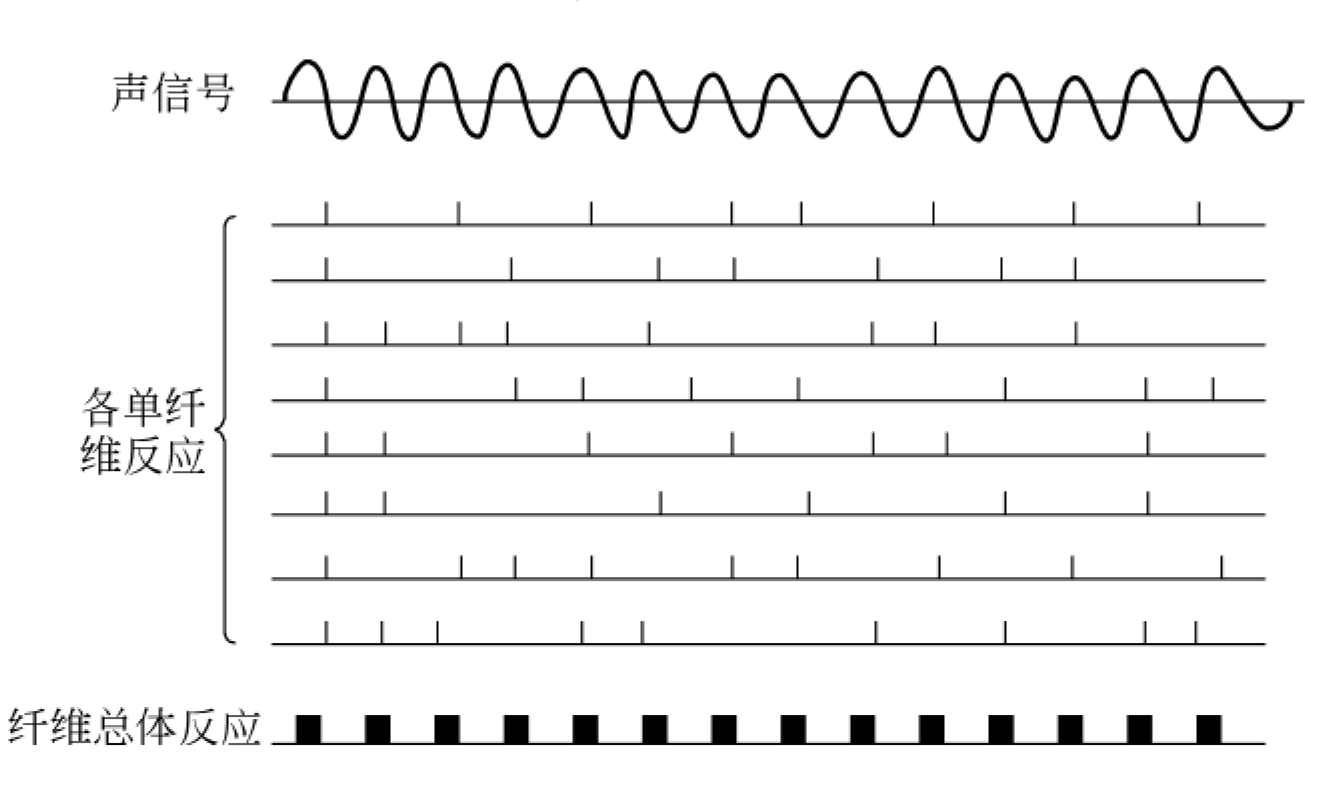



图6.3 排放的组成

排放论认为听神经上冲动排放的频率与声音的频率是一致的,它是频率分析的依据。在听神经纤维上记录神经冲动的实验表明,在声音频率较高时,单根纤维上的神经冲动虽不是每一声波周期都发放一次,但它与声波周期的一定相位总是保持严格的同步关系,称锁相关系。这是排放论很有力的实验证据。

### 3. 频率分析的中枢机理

根据计算耳蜗的频率分辨率是不高的,在1000Hz左右时约为30Hz,而人在1000Hz左右时却可辨别3Hz甚至更小的频率差。精确的频率分析显然是在中枢进行的。在动物从

耳蜗核至听皮层的各级中枢的某些部位，神经细胞的排列都或多或少地有频率区域分布。如同视网膜的各部分在视皮层都有其对应的投射区似的，沿着耳蜗基膜的各部分在听觉皮层表面也有其系统性的特殊投射区。神经单元的放电与刺激声的同步锁相关系在内膝体及它以下各级听觉中枢中都被观察到，在耳蜗核和上橄榄核水平尤为明显。这些事实说明部位原则和时间原则在听觉中枢的频率分析机理中也都是重要的。

心理物理的实验结果表明，人的频率辨别精确度与声音的信号时程之间有特定的函数关系；当信号时程短于某一临界时间 $T$ 时，辨别的精确度与信号时程 $t$ 的平方根成正比，或辨别阈 $\Delta F$ 与 $t$ 的平方根成反比，即 $\Delta F=Kt^{-1/2}$ $(t<T)$。超出这一临界时间后，辨别精确度便恒定在最佳水平，不再随时间的增长而改变，即 $\Delta F=K$ $(t>T)$。$T$ 为 120～150ms。这揭示听觉中枢可能有积累传入信息并对它进行统计学处理的过程，因为若对某一量进行多次测量取其统计值，则测量结果的精确度正是与测量次数的平方根成正比。

## 6.2.2　强度分析机理

声音的强度分析相对地研究得较少。按照一般的规律，感受细胞和神经单元的兴奋阈值有高有低，刺激强时被兴奋的感受细胞和神经单元便多，每一神经单元兴奋后发放神经冲动的数目也大。被兴奋的单元是高阈值的还是低阈值的，兴奋单元的总数是多是少，发放的神经冲动是多是少，这三者都可以成为强度分析的依据。在动物的听神经和听觉中枢，声音刺激强度与反应的上述关系都不难找到实验证据。但有资料表明，听皮层及其他听觉中枢与声音的强度间还有一定的区域分布关系，即不同的区域分别对不同的刺激强度敏感，似乎部位原则也在起作用。

## 6.2.3　声源定位和双耳听觉

声源定位指听觉系统对声源方位的判断，它的基础是双耳听觉。由于从声源到两耳的距离不同及声音传波途中屏障条件的不同，从某一方位发出的声音到达两耳时便有时间差和强度差，它们的大小与声源的方位有关。双耳感受到的声音时间差和强度差便是声源定位的主要依据。对于高频声，强度差的作用较重要，对于低频声，时间差的作用较重要。

除了声源定位这一重要功能外，和单耳听觉相比，双耳听觉有明显的优点。由于双耳综合的结果，声音的响度可以增加，相当于单耳时提高 3～6dB；双耳听觉的辨别能力比单耳好；特别是在有噪声干扰的情况下，双耳听觉对语言的识别能力明显地比单耳高。在双耳听觉的条件下，右耳对语言信号的感受似占较重要的地位，左耳则似对非语言信号的感受较重要，这可能和大脑两半球的分工有关。

## 6.2.4　对复杂声的分析

关于听觉系统如何辨别复杂的问题，目前存在着两种截然不同的观点：①复杂声的感受以听觉系统对其简单组成成分的感受为基础，复杂声在听觉中枢引起的神经活动过程，是各组成成分引起的神经活动过程的总和；②听觉系统有分工检测各种复杂声音或声音某种特征的专门结构单元，称为探测器或特征探测器，它们只对特定的声音或特定的声音特征敏感，对其他声音或声音特征则无反应。两种意见谁是谁非，目前未有权威性定论。

## 6.3 语音编码

语音数字化的技术基本可以分为两大类：第一类方法是在尽可能遵循波形的前提下，将模拟波形进行数字化编码；第二类方法是对模拟波形进行一定处理，但仅对语音和收听过程中能够听到的语音进行编码。其中语音编码的三种最常用的技术是脉冲编码调制(PCM)、差分PCM(DPCM)和增量调制(DM)。通常，公共交换电话网中的数字电话都采用这三种技术。第二类语音数字化方法主要与用于窄带传输系统或有限容量的数字设备的语音编码器有关。采用该数字化技术的设备一般被称为声码器，声码器技术现在开始展开应用，特别是用于帧中继和IP上的语音。

除压缩编码技术外，人们还应用许多其他节省带宽的技术来减少语音所占带宽，优化网络资源。ATM和帧中继网中的静音抑制技术可将连接中的静音数据消除，但并不影响其他信息数据的发送。语音活动检测(SAD)技术可以用来动态的跟踪噪音电平，并为这个噪音电平设置一个享用的语音检测阈值，这样就使得语音/静音检测器可以动态匹配用户的背景噪声环境，并将静音抑制的可听度降到最小。为了置换掉网络中的音频信号，这些信号不再穿过网络，舒适的背景声音在网络的任一端被集成到信道中，以确保话路两端的语音质量和自然声音的连接。语音编码方法归纳起来可以分成三大类：波形编码、信源编码、混合编码。

### 1. 波形编码

波形编码比较简单，编码前采样定理对模拟语音信号进行量化，然后进行幅度量化，再进行二进制编码。解码器作数/模变换后再由低通滤波器恢复出现原始的模拟语音波形，这就是最简单的脉冲编码调制(PCM)，也称为线性PCM。可以通过非线性量化，前后样值的差分、自适应预测等方法实现数据压缩。波形编码的目标是让解码器恢复出的模拟信号在波形上尽量与编码前原始波形相一致，也即失真要最小。波形编码的方法简单，数码率较高，在64kbit/s至32kbit/s之间音质优良，当数码率低于32kbit/s的时候音质明显降低，16kbit/s时音质非常差。

### 2. 信源编码

信源编码又称为声码器，是根据人声音的发声机理，在编码端对语音信号进行分析，分解成有声音和无声音两部分。声码器每隔一定时间分析一次语音，传送一次分析的编码有/无声和滤波参数。在解码端根据接收的参数再合成声音。声码器编码后的码率可以做得很低，如1.2kbit/s、2.4kbit/s，但是也有其缺点。首先是合成语音质量较差，往往清晰度可以而自然度没有，难于辨认说话人是谁，其次是复杂度比较高。

### 3. 混合编码

混合编码是将波形编码和声码器的原理结合起来，数码率约在4～16kbit/s，音质比较好，最近有个别算法所取得的音质可与波形编码相当，复杂程度介乎与波形编码器和声码器之间。

上述的三大语音编码方案还可以分成许多不同的编码方案。语音编码属性可以分为四类，分别是比特速率、时延、复杂性和质量。比特速率是语音编码很重要的一方面。比特速率的范围可以是从保密的电话通信的 2.4kbit/s 到 64kbit/s 的 G.711PCM 编码和 G.722 宽带(7kHz)语音编码器。

## 6.4 韵律认知

韵律是所有自然口语的共同特征，在言语交流中起着非常重要的作用，它通过对比组合音段信息，使说话者的意图得到更好的表达和理解。对人工合成语言而言，韵律控制模型的完善程度，决定了合成语言的自然度。韵律认知日益受到语言学界和言语工程学界的重视。深入全面的理解自然语言的韵律特征无论对语音学研究，还是对提高语音合成的自然度和识别语音的准确性来说，都是至关重要的。语音流信息包括音段信息和韵律信息。音节等音段信息通过音色来表达，韵律信息则通过韵律特征来表达。

言语研究最初为集中探讨句法和语义加工过程，把韵律搁在了一边。一直到了 20 世纪 60 年代，对韵律的系统研究才开始。

### 6.4.1 韵律特征

韵律特征主要包含 3 个方面：重音、语调和韵律结构(指韵律成分的边界结构)。由于它可以覆盖两个或两个以上音段，所以常被称为超音段特征。韵律结构是一个层级结构，对它的成分有各种划分方法，一般公认有 3 个层级，从小到大依次是韵律词、韵律短语和语调短语[968]。

#### 1. 汉语的重音

(1) 词重音和语句重音。

(2) 语句重音的类型。

(3) 语句重音的位置分布及等级差异。

汉语韵律重音是复杂的。在播音员语言里，大量采用对比手段，包括音高的对比和时长的对比。图 6.4 给出下列语句的音高：

“如果以乐器来对应四季，

春天就应该是小号，夏天是定音鼓，

秋天是大提琴，冬天是长笛。”

注意在四个排比的小句中，句末那些重音主要靠时长加大，也就比小句句首的重音低得多了。四小句句首的四季也是一高一低的，后边的乐器也就有相应的一低一高。阳平在句末特低是播音腔。对比重音经常有音节加长的特征。句重音高音线提高也是正常的。句重音并没有另外一种“低重音”的类型。图 6.4 中上声“小、鼓”的音高不太完整，后边的低音段都没有被检测出来。音高范围是 140～400Hz，对数标度。

时长举例：冬 220，天 240，是 235，长 290，笛 270(单位 ms)。

第一小句的调核在“对”，其他小句的调核都在句末。

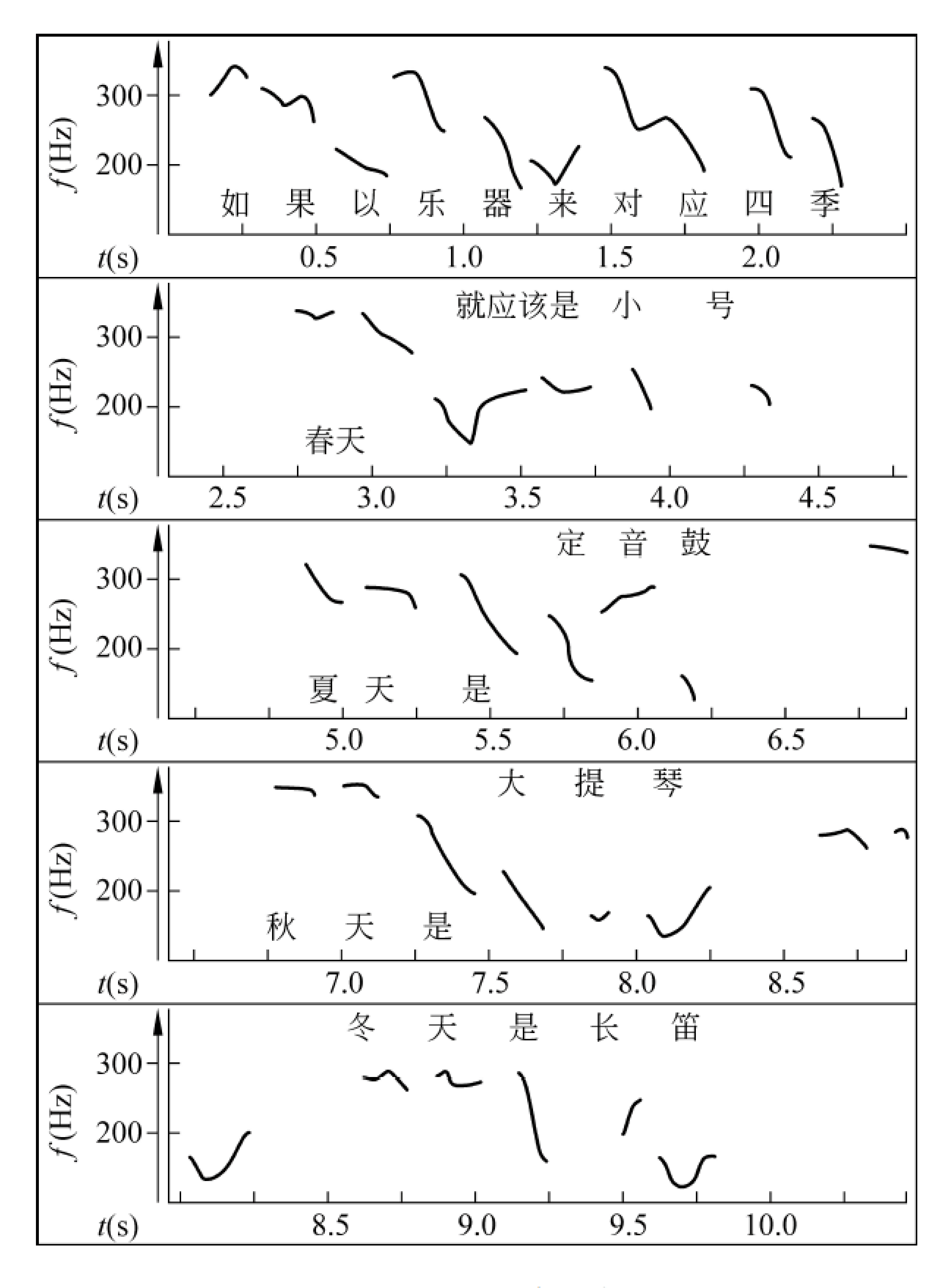



图 6.4 汉语音高的例子

(http://chinese.pku.edu.cn/bbs)

## 2. 汉语的语调

语调构造由语势重音配合而形成。它是一种语音形式，它通过信息聚焦来实施超语法的功能语义。节奏从语言的树型关系出发，按照表达的需要，利用有声形态有限的分解度来安排节奏重音，形成多层套叠的节奏单元。语势重音和节奏重音分别调节声调音域的高音线和低音线。把语调构造各部分的音域特征综合在一起，可以区分出不同的语调类型来，因此语调是声调音域再调节的重要因素，而声调音域是功能语调和口气语调最重要的有声依据。

汉语语调的基本骨架可以分为调冠、调头、调核、调尾四部分。语调构造典型的有声表现如下[910]。

(1) 调头和调核里有较强的语势重音，调冠里只有轻化音节，调尾里一般没有太强的语势重音。

(2) 调核之后声调音域的高音线下移，形成明显的落差。

(3) 调核后一音节明显轻化。调核为上声本调时，后一音节轻化并被明显抬高，高音线落差在其后出现。

大量事实可以说明，在非轻的音节组合中，高音点的高低是跟着语势改变的。在前强后

弱的组合中，后一音节的高音点下移，随着语势差别的扩大，后一个音节还可能轻化。当后边出现较强语势的时候，高音点又可能恢复到一定的音高。

在陈述方式下，高音线下落的过程很快完成，因此调尾一开始或一两个音节后，高音线就已经下移到最低水平。这是高音线的骤降形式。在疑问方式下，高音线下落是逐步完成的，也就是说调尾为首几个音节的高音线呈缓缓下降的曲线形式，这是高音线的渐降形式。高音线和低音线是两种独立的因素，因此陈述语调是高音线骤降形式和低音线下延形式两种特征的组合，不是音域因素单一自由度的调节形式。同样，疑问语调是高音线渐降形式和低音线上敛两种特征的组合。从中分出两个自由度，会有四种基本组合。语音事实表明，除了陈述和疑问外，高音线骤降和低音线上敛是普通的祈使语调，全句高音线渐降和低音线下延是一种重要的感叹语调。四种功能语调还有各种变体，以上描述是很粗的。除了功能语调外，各种口气语调和其他语调现象跟全句或局部的音域调节有关。

语调标记：

（1）高音线无标——骤降语调。调核后高音线骤降，调核后可能出现轻化音节。

（2）高音线有标——渐降语调。调核后高音线渐降，调核后也能出现轻化音节。许多研究都认为，上声调核后出现骤降或渐降的滞后现象。

（3）低音线无标——下延语调。低音线有较大起落，向下延伸。

（4）低音线有标——上敛语调。低音线起落变小，向上收敛，尤以调核和调尾最显著。

通常的疑问语调是强上敛，祈使语调是弱上敛。可能还有其他未知特征。口气语调的分类与功能语调一致，包含相似声学特征，增加带宽特征。它们是高音线和低音线的细微调解。“口气”的意蕴和功能语调的基本功能（语义）相关。

### 3. 韵律结构

（1）韵律词：反映汉语节奏的两音节或三音节组，在韵律词内部不能停顿，在韵律词边界不一定有停顿但是可以有停顿。

（2）韵律词组：一般为两个或三个联系比较紧密的韵律词，在韵律词组内部的韵律词之间通常没有可感知的停顿，而在韵律词组尾一定有一个可以感知的停顿，但从语图上不一定能观察到明显的静音段。

（3）韵律短语：由一个或几个韵律词组组成，韵律短语之间通常有比较明显的停顿，从语图上一般也可以观察到明显的静音段。在音高图上低音线的渐降是韵律短语的重要特点。

（4）语调短语：由一个或几个韵律短语组成。语调短语可以是单句，或复合句中的子句，多为标点符号所隔离。在语调短语后一般有个比较长的停顿。

从上面的定义可以看出，四个韵律单元存在着包含关系，即语调短语边界一定是韵律短语边界，韵律短语边界一定是韵律词组边界，而韵律词组边界只能落在韵律词边界上。但是韵律词组边界不一定是词典词边界，词典词边界也不一定是韵律词边界。

## 6.4.2　韵律建模

人在不同的语境下，会有不同的韵律特征，语境与韵律特征之间具有很强的相关性。韵律特征参数分布受着语境信息的影响，这种影响又满足一定的概率关联关系，而不是一个简

单的函数映射。从概率的角度,对于一个已知的语句参数,与之相对应的韵律特征参数,为所有韵律特征参数中出现概率最大的一组[946],即为:

$$Y = \underset{n}{\operatorname{argmax}} P(Y_n \mid A) \tag{6.1}$$

由贝叶斯公式可以得到

$$Y = \underset{n}{\operatorname{argmax}} P(Y_n \mid A) = \underset{n}{\operatorname{argmax}} \frac{P(A \mid Y_n)P(Y_n)}{P(A)} \tag{6.2}$$

由于 $P(A)$表示语境信息的统计分布,可视其为常数,将其忽略,式(6.2)将进一步转换为:

$$Y = \underset{n}{\operatorname{argmax}} P(Y_n \mid A) = \underset{n}{\operatorname{argmax}} P(A \mid Y_n)P(Y_n) \tag{6.3}$$

式(6.3)表明,为求 $P(Y_n|A)$这样的后验概率,被转换为求 $P(A|Y_n)$这个先验概率,而 $P(Y_n)$则体现了韵律特征本身的分布情况,而这种分布又通过韵律特征间出现概率和相互作用来体现。因而从公式(6.3)中体现了,韵律特征不仅受语境信息的影响,同时韵律特征本身也存在着相互影响的情况。$P(A|Y_n)$作为先验概率的组成部分,从理论上讲,只要是具有数值训练并具有记忆能力的模型,均可以作为其实现的基础。

### 6.4.3 韵律标注

韵律标注是对语音信号中具有语言学功能的韵律特征进行的定性描写。标注语句中有语言学功能的声调变化、语调模式、重音模式和韵律结构,受轻重音影响的音高变化属于标音内容,而元音内在音高变化和音节间声调协同发音不属于标注内容。不标注宜于定量描写的韵律现象。韵律标注一般是分层的,音段切分是韵律标注的基础,所以是必不可少的一层,其他层次的标注,要依据实际应用的需求和标注的语音特性确定。汉语韵律标注系统应该具有如下特点。

(1) 可靠性:不同标音人标音结果的一致性至少要高。

(2) 全面性:覆盖自然语言中最重要的韵律现象。

(3) 易学性:要在短时间内能学会。

(4) 兼容性:要跟语音合成、语音识别的最新方法和当前的句法学、语义学、语用学理论相结合。

(5) 可操作性:标注符号要尽量简单,跟语音表层形式尽量接近。

(6) 开放性:对没有把握的标注项目允许不确定性存在。

(7) 可读性:标注符号机器能认。

从工程应用的角度看韵律标注是对语音进行音系的描写,言语工程通过标注的结果,很容易对语言学信息或语音信息的对应关系进行建模。所以,韵律标注在语音口语识别和基于数据驱动的语音合成系统中,起着越来越大的作用。

韵律标音系统是分音层的,每个音层标记不同的韵律或与之相关的现象。一般来说可以有以下几层信息,使用者可以依据个人需要,选择标注层级。在韵律标注系统 C-ToBI 中分为如下三层。

(1) 音节层:标记普通话音节的拼音形式,如普通话用 1、2、3、4 分别表示四个声调,轻声用 0 表示。声调标在拼音之后。

(2) 实际发音标注层:标注声母、韵母和声调的实际发音。我们采用 IPA 的机读符号

系统 SAMPA-C 标注。

（3）语调层：语调构造是由韵律结构和重音结构决定的，即重音结构和韵律结构确定之后，语调曲线就可以确定了。研究表明，语调变化主要是声调调域（Range）和调阶（Register）的改变，声调变化主要是音高特征的改变。调域受说话人的心理状态、语气、韵律结构等因素的影响会发生改变，所以语调标注应该能够反映调域的扩大和缩小以及调阶的变化和整个语调曲线的变化趋势或叫高低音线的变化趋势。

音段标注就是把话语中的每个语音单元（包括音节、声韵乃至更小的语音单元）逐一进行切分，然后对它们的音色特征分别给予细致如实的描写。汉语音段标注可以分层次进行。音段的扩展标注主要是在正则发音的基础上对语音的实际发音进行标注，将音段和超音段上的音变现象标注出来。所以采用与 IPA 对应的汉语机读音段标注系统 SAMPA-C。对于口语语料，还应该标记副语言学和非语言学现象。

SAMPA 是目前国际上通行的可机读语音键盘符号系统，在语料库的音段标注中广泛使用。在汉语普通话的语音音段标注中已制定出一套可行的 SAMPA-C 符号系统。这里希望将这个系统扩大，包括汉语各地方言。因此首先要确定建立 SAMPA-C 的原则。

（1）依据 SAMPA 符号系统，制定汉语相应的符号。SAMPA（http://www.phon.ucl.ac.uk/home/sampa/）

（2）对汉语特殊语音现象增加附加符号。

（3）一个开放性的系统，随着对语音现象的认识提高，可以增加新的符号或修订不适用的符号。

SAMPA-C 主要是对汉语进行音段标注，首先给出了汉语的辅音元音、声调和音变等标注符号系统。然后把普通话的符号系统列举出来，另外对广州方言、上海方言和福州方言这几个主要方言也给出声韵调符号系统，其他方言可以按照这个模式不断补充进来。

## 6.4.4　韵律生成

韵律生成一开始是作为单词产生的音韵编码过程的一部分受到关注的。随着研究手段的发展，短语和句子产生过程中的韵律生成也得到了研究。这些研究主要是从信息加工的角度进行的。韵律产生的相关模型有 Shattuck-Hufnagel 的扫描复制模型、Dell 的联结主义模型，这两个模型没有专门论述韵律的产生。迄今为止最全面的韵律产生模型是由莱弗特（W J M Levelt）等人提出来的韵律编码和加工模型[416]。

莱弗特认为口语句子的产生过程中，所有阶段的加工都是并行的、递增的。韵律编码包括许多过程，一些在词的范畴进行加工，另一些在句子的范畴进行加工。在一个句子的句法结构展开的同时，词汇的语音规划也产生了。词汇通常分成两部分，Lemma（包含语义和句法特征）的提取和 Lexeme（包含词形及音韵形式）的提取。后者由词形-韵律提取阶段执行，它用 lemma 作为输入来提取相应的词形和韵律结构。所以韵律特征的生成不需要知道音段信息。这些词形和韵律信息被用在音段提取阶段提取词的音段内容（词所包含的音素及其在音节中的位置），然后韵律和音段二者结合在一起。

在最后一个阶段，韵律产生器执行话语语音规划，产生句子的韵律和语调模式。其中韵律的产生包括两个主要步骤：①产生韵律词、韵律短语和语调短语等韵律单元。②产生韵律结构的节律栅。最后用节律仍未表示重音和时间模式。第一步即韵律单元的产生是这样

进行的:词形-韵律提取阶段的加工结果与连接成分组合,成为韵律词。通过扫描句子句法结构,再综合各种相关信息,然后把语法短语的扩展成分包含进来,组成一个韵律短语。而说话者在语流某个点上的停顿,产生语调短语。第二步,在句子韵律结构和单个词的节律栅的基础上,韵律产生器最终构建出整个话语的节律栅。

1999年莱弗特又提出了单词的产生过程中韵律生成的新观点[418],认为在荷兰语、英语和德语这样的重音语言中,存在一个主要的词韵律模式:即词重音放在第一个全元音音节上。所以规则词的重音在递增的音节化过程中是遵照这个规则自动产生的,而不是提取的。不规则词的重音不能自动产生。所以,只有不规则词的韵律结构,才作为音韵代码的一部分被储存起来。不规则词的韵律结构提取出来后,被用来指导不规则词的韵律化。韵律化的结果是产生音节和更大的韵律单元。

这些模型主要是建立在英语、荷兰语和德语的研究成果基础上的。和这些语言相比,汉语语音有两个显著特点:①音节数量少,只有英语等语言的十分之一左右;②汉语是一种声调语言,而英语等语言是重音语言,没有声调。所以汉语韵律产生机制必然和这些模型有不同之处[968]。不过目前对汉语韵律产生的研究还不是很多。

### 6.4.5 韵律生成的认知神经科学机制

莱弗特等用元分析法分析了58个脑功能成像研究结果[425],总结词汇产生过程中,脑区的激活呈左侧化趋势,包括后额下回(Broca区)、颞上回中部、颞中回、后颞上回、后颞中回(Wernicke区)和左丘脑。视觉和概念上的引入过程涉及枕叶、腹侧颞叶和额前区(0~275ms);接着激活传至Wernicke区,单词的音韵代码存储在该区,这种信息传播至Broca区和(或)颞左中上叶,进行后音韵编码(275~400ms);然后进行语音编码,这一过程与感觉运动区和小脑有关,激活感觉运动区进行发音(400~600ms)。

2002年梅耶(Mayer)等人[469]用fMRI研究正常人韵律产生过程中的大脑活动,发现左右半球的前头骨-前头盖底的相对较小且不重叠的区域与韵律产生有关。语言学韵律的产生仅激活左半球,而情感韵律的产生则仅仅激活右半球。

## 6.5 语音识别

自动语音识别(Automatic Speech Recognition,ASR)是实现人机交互尤为关键的技术,让计算机能够“听懂”人类的语音,将语音转化为文本。自动语音识别技术经过几十年的发展已经取得了显著的成效。近年来,越来越多的语音识别智能软件和应用走入了大家的日常生活,苹果的Siri、微软的小娜(Cortana)、百度度秘(Duer)、科大讯飞的语音输入法和灵犀等都是其中的典型代表。随着识别技术及计算机性能的不断进步,语音识别技术在未来社会中必将拥有更为广阔的前景。

### 6.5.1 语音识别概况

以1952年贝尔实验室研制的特定说话人孤立词数字识别系统为起点,语音识别技术已经历了60多年的持续发展。其发展历程可大致分为以下四个阶段。

### 1. 20 世纪 50 年代至 70 年代

该阶段是语音识别的初级阶段，主要研究孤立词识别。在动态时间规整技术、线性预测编码技术、矢量量化技术等取得进展。IBM 公司的杰利内克(F Jelinek)等在 20 世纪 70 年代末提出 n-gram 统计语言模型，并成功地将 trigram 模型应用于 TANGORA 语音识别系统中。此后美国卡内基梅隆大学采用 bigram 模型应用于 SPHINX 语音识别系统，大大提高了识别率。此后一些著名的语音识别系统也相继采用 bigram，trigram 统计语言模型用于语音识别系统。

### 2. 20 世纪 80 年代至 90 年代中期

识别算法从模式匹配技术转向基于统计模型的技术，更多地追求从整体统计的角度来建立最佳的话音识别系统。最典型的为隐马尔科夫模型(Hidden Markov Model，HMM)在大词汇量连续语音识别系统中的成功应用。美国国防部先进研究项目局(Defense Advanced Research Projects，DARPA)自 1983 年开始为期十年的 DARPA 战略计算工程项目，其中包括用于军事领域的语音识别和语言理解、通用语料库等。参加单位包括 MIT(麻省理工学院)、CMU(卡内基梅隆大学)、BellLab 和 IBM 公司等。20 世纪 80 年代末，美国卡耐基梅隆大学用 VQ-HMM 实现了语音识别系统 SPHINX，这是世界上第一个高性能的非特定人、大词汇量、连续语音识别系统，开创了语音识别的新时代。至 90 年代中期，语音识别技术进一步成熟，并出现了一些很好的产品。该阶段可以认为是统计语音识别技术的快速发展阶段。

### 3. 20 世纪 90 年代中期至本世纪初

该阶段语音识别研究工作更趋于解决在真实环境应用时所面临的实际问题。美国国家标准技术局和美国国防部先进研究项目局组织了大量的语音识别技术评测，极大地推动了该技术的发展。在此阶段，基于高斯混合模型(Gaussian Mixture Model，GMM)和 HMM 的混合语音识别框架成为领域内的主流技术。而区分度训练技术的提出，进一步提升了系统性能。此外，为提升系统的鲁棒性及实用性，语音抗噪技术、说话人自适应训练(Speaker Adaptive Training，SAT)等技术被相继提出。该阶段可看作是 GMM-HMM 混合语音识别技术趋于成熟并应用的阶段。

### 4. 21 世纪初至今

该阶段的特点为基于深度学习的话音识别技术成为主流。以 2011 年提出的上下文相关-深度神经网络-隐马尔科夫框架为变革开始的标志。基于链接时序分类(Connectionist Temporal Classification，CTC)[282]搭建过程简单，且在某些情况下性能更好。2016 年，谷歌提出 CD-CTC-SMBR-LSTM-RNNS，标志着传统的 GMM-HMM 框架被完全替代。声学建模由传统的基于短时平稳假设的分段建模方法变革到基于不定长序列的直接判别式区分的建模。由此，语音识别性能逐渐接近实用水平，而移动互联网的发展同时带来了对语音识别技术的巨大需求，两者相互促进。与深度学习相关的参数学习算法、模型结构、并行训练平台等成为该阶段的研究热点。该阶段可看作是深度学习语音识别技术高速发展并大规模

应用的阶段。

我国语音识别研究工作起步于 20 世纪 50 年代,而研究热潮是从 20 世纪 80 年代中期开始。在 863 计划的支持下,中国开始了有组织的语音识别技术的研究。语音识别正逐步成为信息技术中人机接口的关键技术,研究水平也从实验室逐步走向实用。

## 6.5.2 语音识别系统结构

语音识别系统包含四个主要模块:前端信号处理与特征抽取、声学模型、语言模型以及解码器(见图 6.5)。

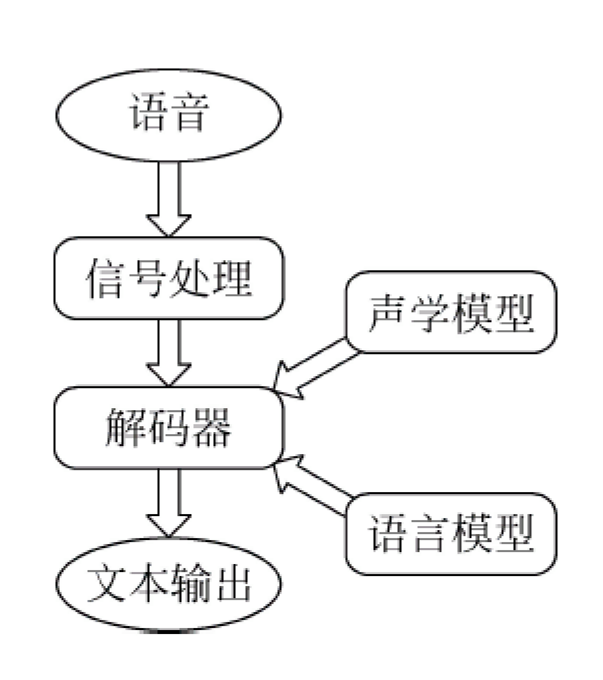



图 6.5 语音识别系统框架

信号处理模块输入为语音信号,输出为特征向量,随着远场语音交互需求越来越大,前端信号处理与特征提取在语音识别中位置越来越重要。一般而言,主要过程为首先通过麦克风阵列进行声源定位,然后消除噪声。通过自动增益控制将收音器采集到的声音放到正常幅值。通过去噪等方法对语音进行增强,然后将信号由时域转换到频域,最后提取适用于 AM 建模的特征向量。

声学模型对声学和发音学知识进行建模,其输入为特征抽取模块产生的特征向量,输出为某条语音的声学模型得分。声学模型是对声学、语音学、环境的变量,以及说话人性别、口音的差异等的知识表示。声学模型的好坏直接决定整个语音识别系统的性能。

语言模型则是对一组字序列构成的知识表示,用于估计某条文本语句产生的概率,称为语言模型得分。模型中存储的是不同单词之间的共现概率,一般通过从文本格式的语料库中估计得到。语言模型与应用领域和任务密切相关,当这些信息已知时,语言模型得分更加精确。

解码器根据声学模型和语言模型,将输入的语音特征矢量序列转化为字符序列。解码器将所有候选句子的声学模型得分和语言模型得分融合在一起,输出得分最高的句子作为最终的识别结果。

## 6.5.3 基于深度神经网络的语音识别系统

基于深度神经网络的语音识别系统框架如图 6.6 所示。相比传统的基于 GMM-HMM 的语音识别系统,其最大的改变是采用深度神经网络替换 GMM 模型对语音的观察概率进行建模。最初主流的深度神经网络是最简单的前馈型深度神经网络(Feedforward Deep Neural Network,FDNN)。FDNN 相比 GMM 的优势在于:①使用 FDNN 估计 HMM 的状态的后验概率分布不需要对语音数据分布进行假设;②FDNN 的输入特征可以是多种特征的融合,包括离散或者连续的;③FDNN 可以利用相邻的语音帧所包含的结构信息。

考虑到语音信号的长时相关性,一个自然而然的想法是选用具有更强长时建模能力的神经网络模型。于是,循环神经网络(Recurrent Neural Network,RNN)近年来逐渐替代传统的 FDNN 成为主流的语音识别建模方案。如图 6.7 所示,相比前馈型神经网络 FDNN,循环神经网络在隐层上增加了一个反馈连接,也就是说,RNN 隐层当前时刻的输入有一部

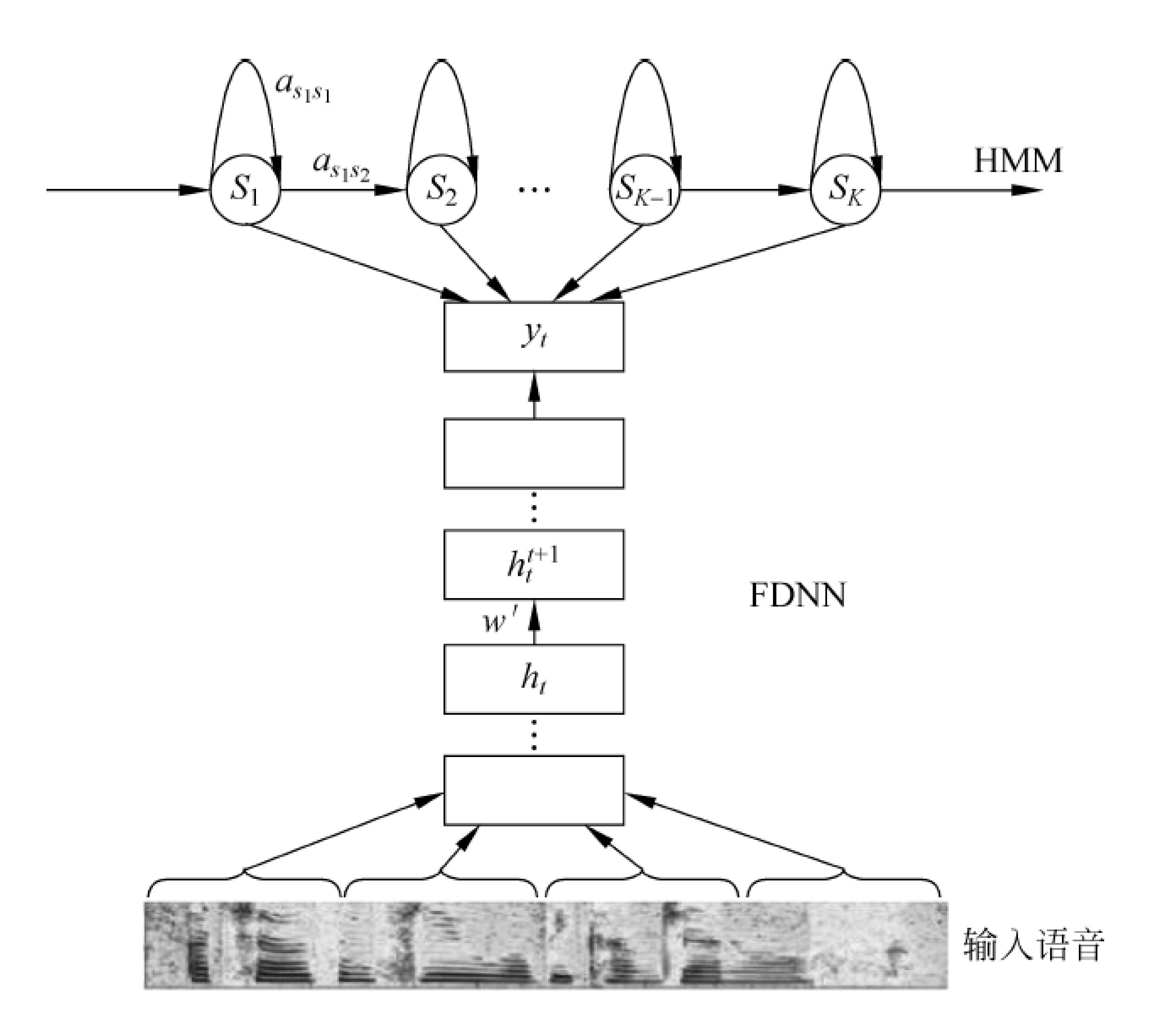



图 6.6 基于深度神经网络的语音识别系统框架

分是前一时刻的隐层输出,这使得 RNN 可以通过循环反馈连接看到前面所有时刻的信息,这赋予了 RNN 记忆功能。这些特点使得 RNN 非常适合用于对时序信号的建模。而长短时记忆模块(Long-Short Term Memory,LSTM)的引入解决了传统简单 RNN 梯度消失等问题,使得 RNN 框架可以在语音识别领域实用化并获得了超越 FDNN 的效果,目前已经使用在业界一些比较先进的语音系统中。除此之外,研究人员还在 RNN 的基础上做了进一步改进工作,如图 6.8 是当前语音识别中的主流 RNN 声学模型框架,主要包含两部分:深层双向 RNN 和链接时序分类 CTC 输出层。其中双向 RNN 对当前语音帧进行判断时,不仅可以利用历史的语音信息,还可以利用未来的语音信息,从而进行更加准确的决策;CTC 使得训练过程无须帧级别的标注,实现有效的“端对端”训练。

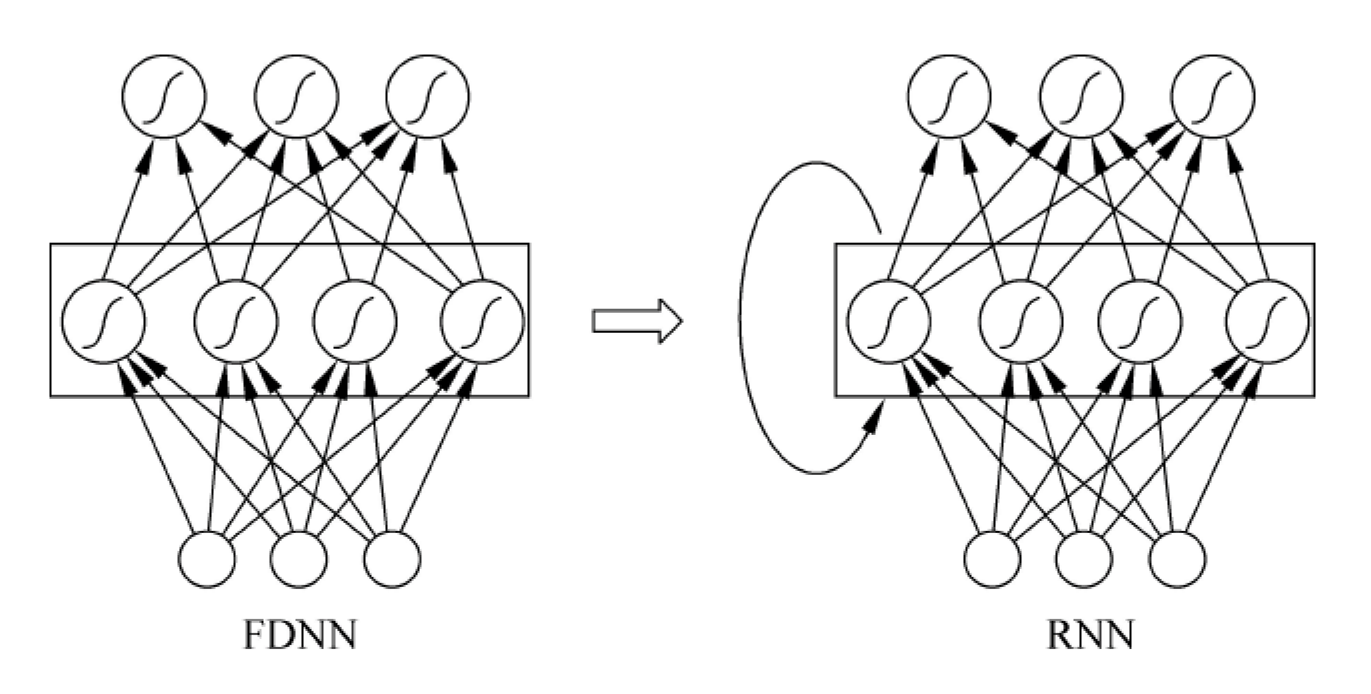



图 6.7 FDNN 和 RNN 示意图

语音识别任务是将输入波形映射到最终的词序列或中间的音素序列。声学模型真正应该关心的是输出的词或音素序列,而不是在传统的交叉熵训练中优化的一帧一帧的标注。为了应用这种观点并将语音输入帧映射成输出标签序列,链接时序分类 CTC 方法被引入了进来。为了解决语音识别任务中输出标签数量少于输入语音帧数量的问题,链接时序分类 CTC 引入

了一种特殊的空白标签,并且允许标签重复,从而迫使输出和输入序列的长度相同。

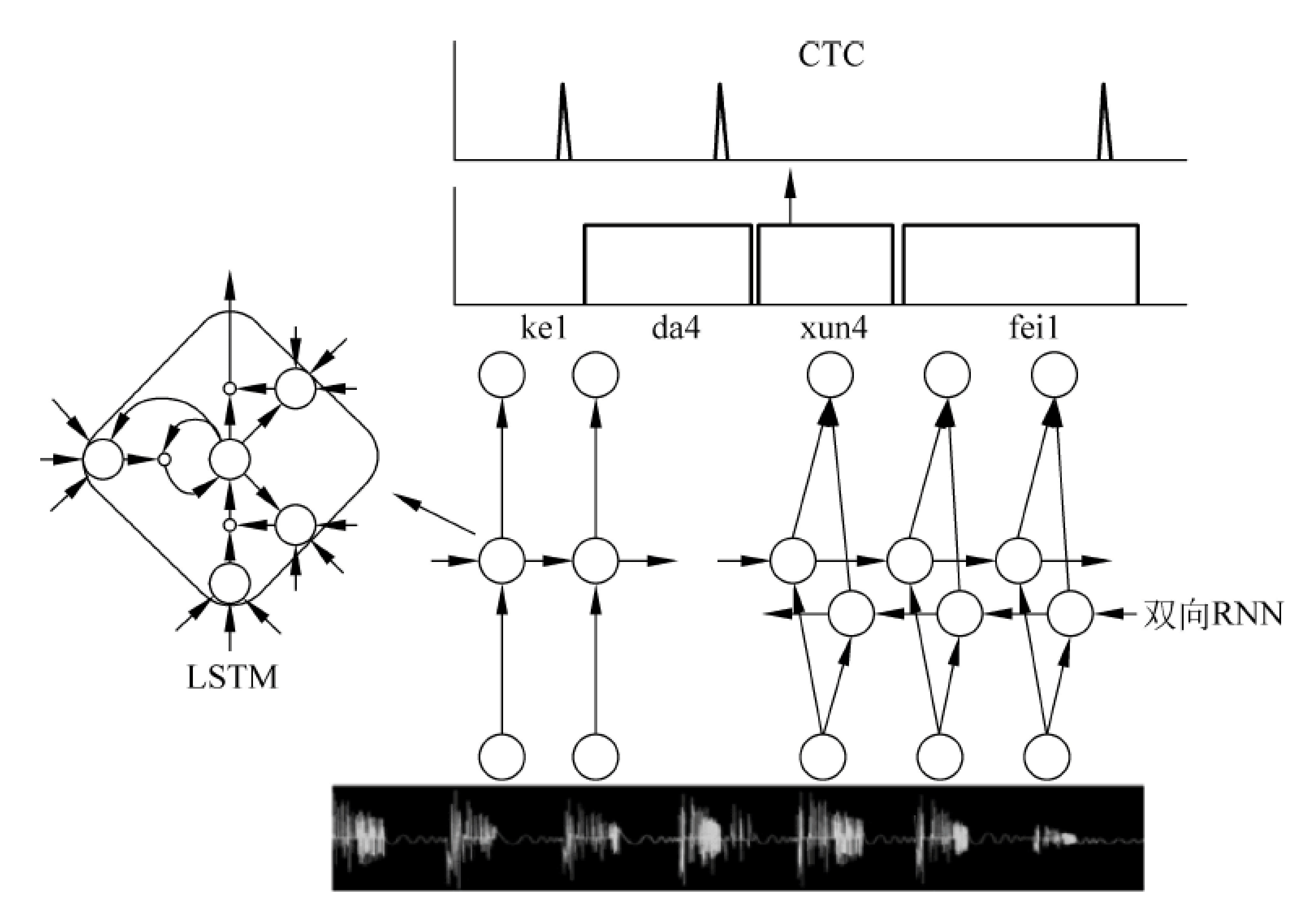



图 6.8 基于 RNN—CTC 的主流语音识别系统框架

链接时序分类 CTC 的一个迷人特点是我们可以选择大于音素的输出单元,例如音节和词。这说明输入特征可以使用大于 10ms 的采样率构建。链接时序分类 CTC 提供了一种以端到端的方式优化声学模型的途径。用端到端的语音识别系统直接预测字符而非音素,从而也就不再需要使用词典和决策树了。

早在 2012 年,卷积神经网络 CNN 就被用于语音识别系统,但始终没有大的突破。最主要的原因是没有突破传统前馈神经网络采用固定长度的帧拼接作为输入的思维定式,从而无法看到足够长的语音上下文信息。另外一个缺陷是只将 CNN 视作一种特征提取器,因此所用的卷积层数很少,一般只有一到二层,这样的卷积网络表达能力十分有限。

讯飞研发了深度全序列卷积神经网络(Deep Fully Convolutional Neural Network, DFCNN)的语音识别框架,使用大量的卷积层直接对整句语音信号进行建模,更好地表达了语音的长时相关性。DFCNN 的结构如图 6.9 所示,它直接将一句语音转化成一张图像作为输入,即先对每帧语音进行傅里叶变换,再将时间和频率作为图像的两个维度,然后通过非常多的卷积层和池化(Pooling)层的组合,对整句语音进行建模,输出单元直接与最终的识别结果比如音节或者汉字相对应。

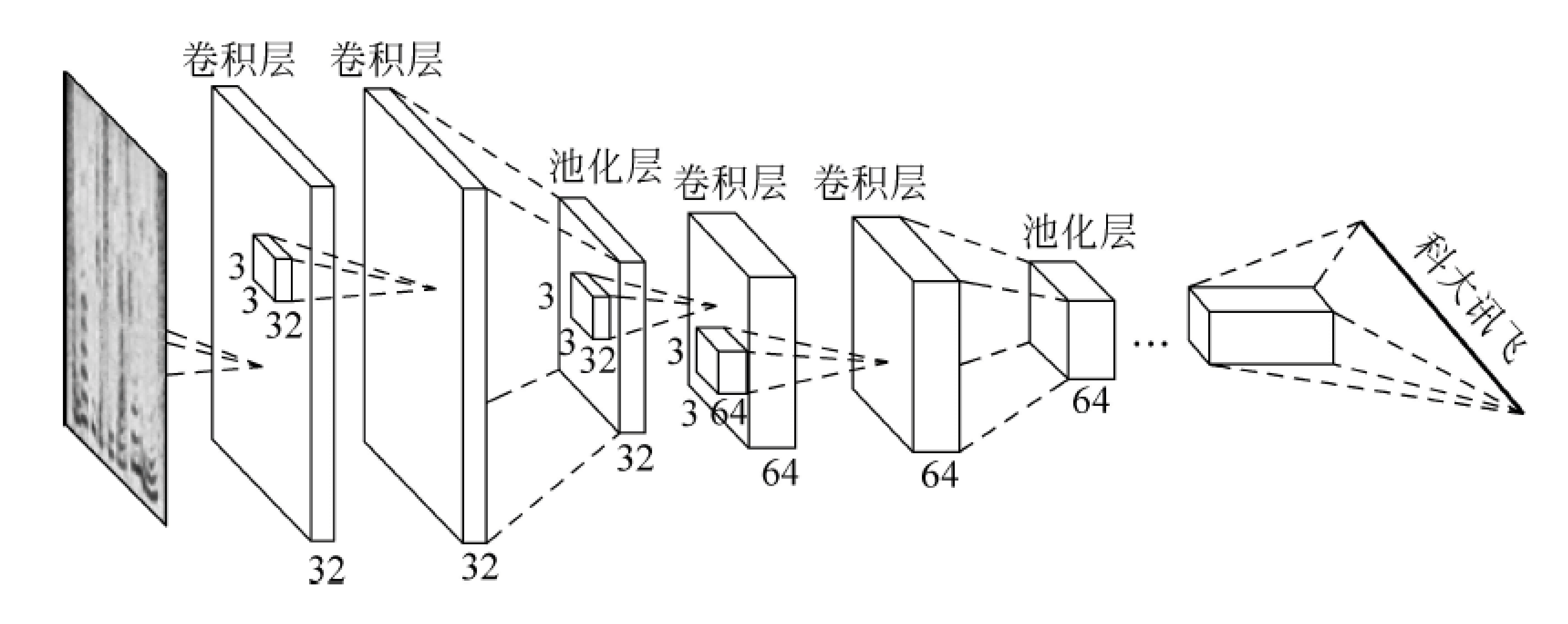



图 6.9 DFCNN 示意图

DFCNN 直接将语谱图作为输入，与其他以传统语音特征作为输入的语音识别框架相比具有天然的优势。从模型结构来看，DFCNN 与传统语音识别中的 CNN 做法不同，它借鉴了图像识别中效果最好的网络配置，每个卷积层使用 3×3 的小卷积核，并在多个卷积层之后再加上池化层，这样大大增强了 CNN 的表达能力，与此同时，通过累积非常多的这种卷积池化层对，DFCNN 可以看到非常长的历史和未来信息，这就保证了 DFCNN 可以出色地表达语音的长时相关性，相比 RNN 网络结构在鲁棒性上更加出色。最后，从输出端来看，DFCNN 还可以和近期很热的 CTC 方案完美结合以实现整个模型的端到端训练，且其包含的池化层等特殊结构可以使得以上端到端训练变得更加稳定。

在和其他多个技术点结合后，科大讯飞 DFCNN 的语音识别框架在内部数千小时的中文语音短信听写任务上，相比目前业界最好的语音识别框架双向 RNN-CTC 系统获得了 15% 的性能提升，同时结合科大讯飞的 HPC 平台和多 GPU 并行加速技术，训练速度也优于传统的双向 RNN-CTC 系统。DFCNN 的提出开辟了语音识别的一片新天地。

## 6.6　语音合成

语音合成即让计算机生成语音的技术，其目标是让计算机能输出清晰、自然、流畅的语音。按照人类言语功能的不同层次，语音合成也可以分成三个层次，即从文字到语音的合成、从概念到语音的合成、从意向到语音的合成。这三个层次反映了人类大脑中形成说话内容的不同过程，涉及人类大脑的高级神经活动。目前的成熟的语音合成技术只能够完成从文字到语音（Text-to-Speech，TTS）的合成，该技术也常常被称作文语转换技术。

### 6.6.1　语音合成概况

语音合成技术的研究已有两百多年的历史，但真正具有实用意义的近代语音合成技术是随着计算机技术和数字信号处理技术的发展而发展起来的，主要是让计算机能够产生高清晰度、高自然度的连续语音。在语音合成技术的发展过程中，早期的研究主要集中在言语产生的机理、语音的特征、发音器官的模拟等方面。20 世纪 80 年代末，基于时域波形拼接方法有了较大的发展。随着 20 世纪 90 年代计算机运算和存储能力的飞速发展，基于大语料库的基元选取和波形拼接方法逐渐成为语音合成的主流方法，并且获得了非常成功的商业应用。

20 世纪末，随着语音信号统计建模方法的日渐成熟以及用户对语音合成系统构建自动化程度需求的提高，基于统计建模的语音合成方法被提出。这种方法的思想是采用统计的方法对语音信号建模，用机器学习的方法建立文本特征与韵律和声学参数之间的映射关系，其中尤以基于隐马尔科夫模型[618]的语音合成方法为代表。该方法可以自动快速地构建合成语音，并且对不同的发音人、不同的发音风格、甚至语种的依赖性非常小，而且系统资源占用很小，合成语音的音色和风格易于控制，因此该方法一度成为与基于语料库的基元选取和拼接合成相并列，是一种主流语音合成方法，时至今日仍有很多的工业级合成系统采用该方法。

近年来，随着计算机计算性能的逐步提高，以及深度学习技术的迅猛发展，基于大数据

的深度学习算法逐渐成为语音合成的主流方法。神经网络建模逐渐替代了基于 HMM 的建模方法。基于神经网络的语音合成是以神经网络为基础,完成从文本参数到声学参数映射的模型框架。这一框架的应用是人工智能时代和大数据给语音合成带来的最重要的改变。深度学习的技术显著提高了语音合成的音质、自然度以及情感丰富程度,同时也降低了语音合成开发的专业门槛,使得人们不再需要精通语音语言学、自然语言处理、音韵学、语音信号处理乃至心理学的各个领域知识来设计一个复杂的语音合成系统,而是从大语料中通过统计学习方法得到语音合成的规律,进而实现高质量乃至工业级的语音合成系统。

## 6.6.2 文字到语音合成系统

典型的文字到语音合成系统如图 6.10 所示,该系统可以分为文本分析模块、韵律预测模块和声学模型,下面对三个模块进行简要的介绍。

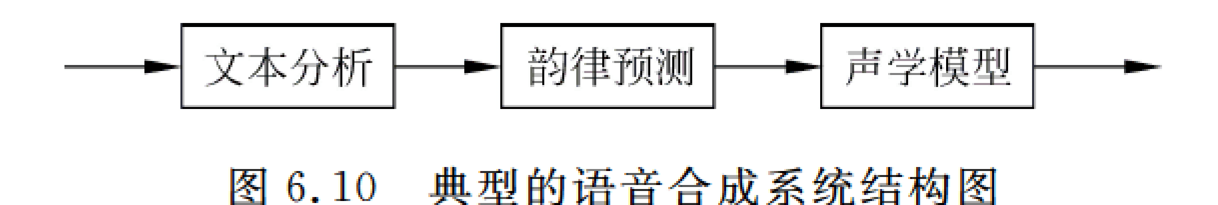



图 6.10 典型的语音合成系统结构图

### 1. 文本分析模块

文本分析模块是语音合成系统的前端。它的作用是对输入的任意自然语言文本进行分析,输出尽可能多语言相关的特征和信息,为后续的系统提供必要的信息。它的处理流程依次为:文本预处理、文本规范化、自动分词、词性标注、字音转换、多音字消歧、字形到音素(Grapheme to Phoneme, G2P)、短语分析等。文本预处理包括删除无效符号、断句等。其中,文本规范化的任务就是将文本中的非普通文字(如数学符号、物理符号等)字符识别出来,并转化为一种规范化的表达。字音转换的任务是将待合成的文字序列转换为对应的拼音序列。多音字消歧则是解决一字多音的问题。G2P 是为了处理文本中可能出现的未知读音的字词,这在英文或其他单词以字母组成的语言中经常出现。

### 2. 韵律预测模块

韵律即是实际话流中的抑扬顿挫和轻重缓急,例如重音的位置分布及其等级差异,韵律边界的位置分布及其等级差异,语调的基本骨架及其跟声调、节奏和重音的关系等等。由于这些特征需要通过不止一个音段上的特征变化得以实现,通常也称之为超音段特征。韵律表现是一个很复杂的现象,对韵律的研究涉及到语音学、语言学、声学、心理学等多个领域。韵律预测模块则接收文本分析模块的处理结果,预测相应的韵律特征,包括停顿、句重音等超音段特征。韵律模块的主要作用是保证合成语音拥有自然的抑扬顿挫,提高语音的自然度。

### 3. 声学模型模块

声学模型的输入为文本分析模块提供的文本相关特征和韵律预测模块提供的韵律特征,输出为自然语音波形。目前主流的声学模型采用的方法可以概括为两种:一种是基于时域波形的拼接合成方法,声学模型模块首先对基频、时长、能量和节奏等信息建模,并在大规模语料库中根据这些信息挑选最合适的语音单元,然后通过拼接算法生成自然语音波形;

另一种是基于语音参数的合成方法，声学模型模块根据韵律和文本信息的指导来得到语音的声学参数，如谱参数、基频等，然后通过语音参数合成器来生成自然语音波形。

语音合成系统的声学模型从所采用的基本策略来看，可以分为基于发音器官的模型和基于信号的模型两大类。前者试图对人类的整个发音器官进行直接建模，通过该模型进行语音的合成，该方法也被称为基于生理参数的语音合成。后者则是基于语音信号本身进行建模或者直接进行基元选取拼接合成。相比较而言，基于信号模型的方法具有更强的应用价值，因而得到了更多研究者和工业界的关注。基于信号模型的方法有很多，主要包括基于基元选取的拼接合成和统计参数语音合成。

### 6.6.3　概念语音转换系统

#### 1. 系统结构

日本大坂大学实现了一种概念语音(Concept To Speech，CTS)转换系统。这是一个基于格结构表述的语音输出系统 SCOS(http://www.zzrtu.com/jsj/)。

CTS 的输入是建立在格结构和短语模型基础上的概念描述，输出的是合成语音。这一层次的研究包括自然语言生成、韵律修饰和语音合成。它首先要由概念描述生成带有韵律修饰特性的语句，然后将语句转换成语音输出。

对 CTS 系统来说，如何表示系统输入是一个值得研究的重要问题。在 SOCS 中，基于格结构和短语模型之上的表述，即概念表述，被转换成合成语音。这个概念描述是在一个抽象的概念水平上描绘的。它使按照对话内容来修饰输出语句成为可能，同时也有可能构造适应性强的言语输出界面。SOCS 以两种内置的机制来控制概念表述的韵律参数：第一个是停顿标记(Pause Marker)，这是在句子生成时和单词一块生成的。停顿标记用于插入停顿，也可以用于设置韵律短语边界的基频。第二是嵌于惯用模板内的韵律修饰函数 PMF (Prosody Modification Function)。SMF 函数控制句型的韵律参数。

SOCS 可能作为智能工作系统(Intelligent Performance Systems，IPS)的通用接口。以往 IPS 产生文本方式的消息，并在终端上显示。但人们期望，通过语音与 IPS 交换信息。当前语音来自文语转换系统，而获得语音更方便的表示方式不应是文本。SOCS 与 IPS 的接口框图如图 6.11 所示。

SOCS 是把概念表述转换成语音的 CTS 系统。IPS 产生的中间表述形式叫作概念表述，通过对话管理器它被送到 SOCS。对话管理器根据上下文修改概念表述，并产生句子和韵律特征。CTS 的输入是用语义表示框架，而不是用文本来表示的。它比 TTS 具有如下优点：CTS 系统直接由概念表述产生句子，因而没有必要进行文本分析。一些特征如重音、语速可直接加到概念表述中。一种概念表述可以根据上下文转换成几种句子。故此 IPS 只需决定“说什么”，而不用管“怎么说”。因此，CTS 是人机语音交互的一种必要手段。

#### 2. 概念表述

(1) 中间表述的要求

要实现人机之间真正自然的对话，有很多问题需要解决。首先是对话的总体处理：如话语中间的完整和省缺；对人意图的理解；依据情景采用最合适的表达方式等。其次是人

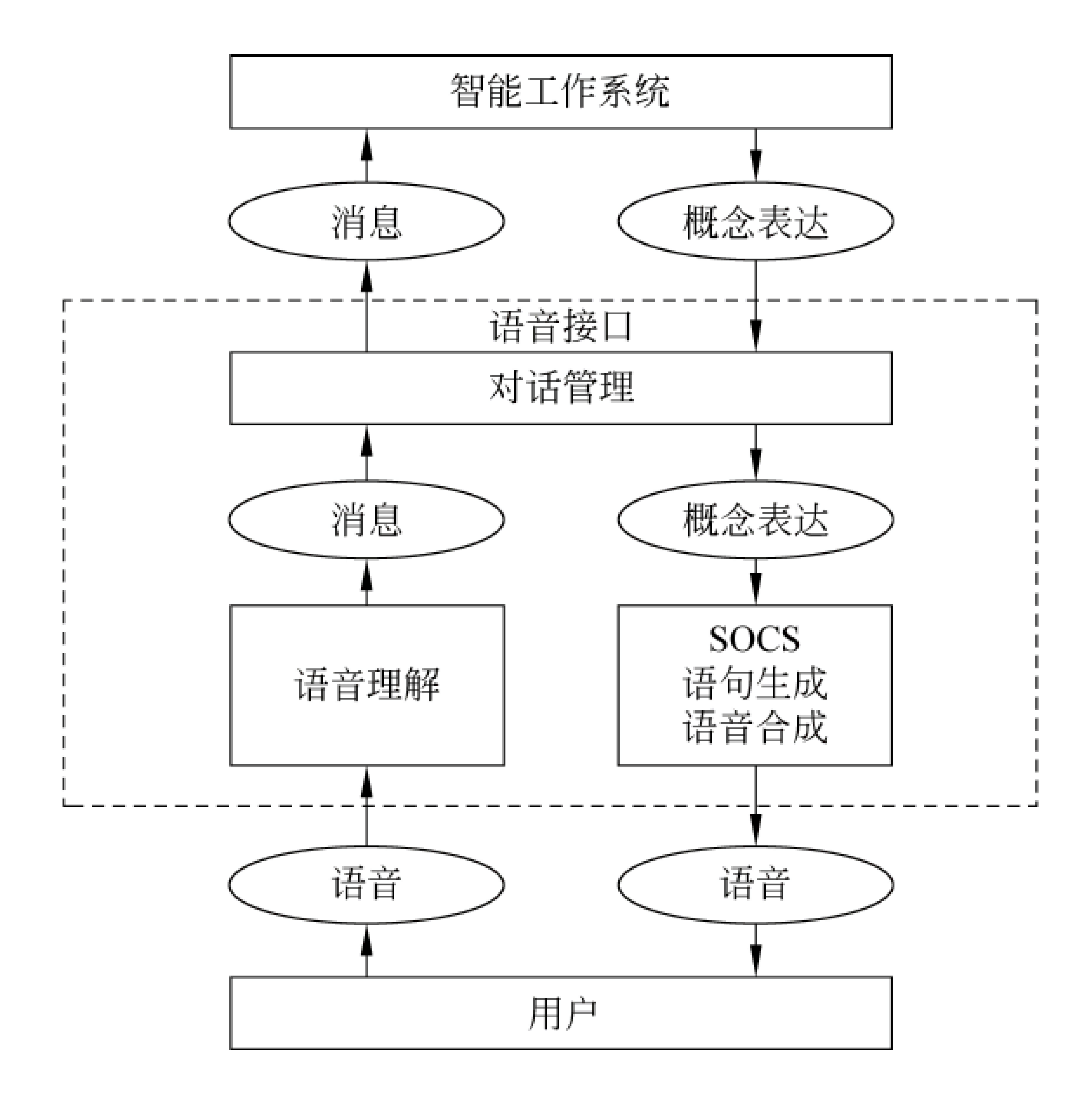



图 6.11　SOCS 与 IPS 的接口

机语言交互的处理：如语音识别错误后的恢复、输出语音的韵律生成等。这些问题不仅有赖于 IPS 中的处理，还源于对话本身。因此应设计一个独立于 IPS、能灵活地处理这些现象的语音界面是必要的。但自然语言复杂、表达能力很强，而且对话过程中，基于上下文的表述不断地变动。在界面中自动修改自然语言语句是相当难的。总之，信息描述的表示方案起着关键的作用。

SOCS 系统试图找到一种比自然语言更有弹性的中间表述方式，它的特点是：

① 易于抽取用于韵律控制的必要信息。期望 CTS 对韵律的控制比 TTS 容易。

② 易于产生语句。CTS 从中间表示生成语句，而 TTS 必须分析文本，因此可能通过简单的机制实现。

③ 易于修改中间表述形式(删除单词、增加韵律特性)，以产生适合于上下文的语句。

④ 当 IPS 执行时，易于准备 IPS 产生中间形式所需的数据。

⑤ 中间形式的表达力应较强，一种表示应能转换成不同的句子和具有不同韵律结构的语言表达。

⑥ 易于 IPS 产生中间形式。

比较了众多表示方式后发现把基于格结构的表述和一些模板结合起来，可满足上述要求。图 6.12 是概念表述的例子。在这种表述方案中，给出了表层语句使用的词语和含义，但没有给出语法结构。概念表述减小了 CTS 生成语句的复杂性，增加了依据对话上下文变化语句的可能性。

(2) 概念表述的组成

这里概念表述的定义是基于格结构的。它由七部分组成。图 6.13 中以“$”开头的串都是 SOCS 系统定义的模板或操作符。如 $why()就是一个惯用模板。

① 原子符号：单词(动词、名词、形容词等)作为原子符号。

```
s1(hajimeru([ $ object(lesson1), $ persuasive])).
  ('star')
lesson1( $ modify(chiri , benkyo , [ ]).
('geography')('lesson')s2( $ because(soviet1,seiikusuru([ $ actor(kome), $ negative]))).
('grow') ('rice') soviet1( $ is(soren, $ modify(samui,kuni,[ $ p_priminent(samui)])).
('the Soviet union') ('cold') ('country') ('cold')
 $ 3( $ why(seiikusuru([ $ actor(kome), $ place(iran)]),[ $ p_interrogative])).
('grow') ('rice') ('Iran')
```

图 6.12　概念表述的例子

② 修饰模板：$ Modify (A,B,OPs)表示"A"与"B"之间的修饰关系("A"和"B"是句子的成分)。

③ 格模板：用来表示动词的格信息。它表示了动作和受主之间的关系。SOCS 中定义了十三种格模板。

④ 语气操作符：描述动词的时态、体态和形式等信息，语气操作符将动词按事先存储的模式变成适当的形式。

⑤ 连接模板：表示两个连续句子之间的关系。

⑥ 惯用模板：用于表达 IPS 输出信息中经常使用的语句/短语模式和不是基于格关系的语义结构。

⑦ 韵律操作符：它与动词和模板一起指明韵律特性。SOCS 有三个韵律操作符：$ p_prominent(A)，$ p_interrogative，$ p_speed(A,S)。

在 SOCS 中，惯用模板对生成韵律模式扮演重要角色。

```
   $ reason(A,B,Ops): [A,wa,p2,B,kara,desu], $ mod_acc(4, + ,15), $ mod_bpit(4, + ,10).
   ('the reason of A is B')
 $ because(A,B,OPs): [A,kara,p2,B], $ mod_bpit(4,190).
   ('because A,B')
 $ why(A,OPs): [naze,p2,A,no,desu,ka], $ mod_dur(1, + ,10), $ mod_bpit(3,190).
   ('why A')
 $ is(A,B,OPs): [A,wa,pl,B,desu].
   ('A is B')
```

图 6.13　惯用模板

惯用模板利用词汇列表产生句子和短语。词汇列表中，除词汇之外，还有停顿标记"p1"和"p2"。停顿符号穿插在形成的句子中，它们将在韵律控制的过程中使用。停顿符号也会用在其他模板中，如修饰模板、格模板和连接模板。韵律修饰函数(PMF)能与惯用模板联系起来，控制韵律参数，如时长、韵律短语起始时的 f0、词语的重音。图 6.13 $ reason 中的 $ mod-acc 和 $ mod-bpit 是 PMF 的例子。

(3) 语句形成

SOCS 是按照模板引用和格语法生成语句的，句子中包括句法结构和停顿标记。动词短语根据格描述，按预定的顺序连接格元素而形成的。

连成语句后，再自动变换成口语的礼貌形式，句尾加-dasu 或-masu。这样，便形成了

IPS 使用者的友好界面。图 6.14 表示了由图 6.12 概念表示生成的语句。图 6.12 中 s1 中的 hajimeru 先变成礼貌形式 hajimemasu,然后 $persuasive 操作符给 masu 后面加上功能词,映射成 macho。这样,hajimeru 的结果是 hajimemashou。

```
 s1: chiri no pl penkyo wo pl hajimemashou
('Let us start studying geography.')
 s2: soren wapl samui kuni desu kara p2 kame ga pl seiikushimasen
('As the Soviet union is a cold country , rice not grow.')
 s3: naze p2 kome ga pl iran de pl seiikusuru no desu ka
('Why does rice grow in Iran?')
```

图 6.14　生成的语句

(4) 韵律控制

① 停顿标记:按照生成的语句中的停顿标记插入停顿。停顿时间有如下两种。

p2:停顿 300ms (PS2)

p1:停顿 100ms (PS1)

如果短语长度太长,超过 25 个字符,可在这个短语中插入一个 PS1,即把该短语分成两个一次性呼吸短语。注意,词和短语标记是在一个网状结构中生成的。若有不止一个 p1 时,则选择最外层的 p1。

② 韵律修改函数(PMF)。PMF 函数用于在惯用模板中改变句子的局部韵律参数。目前有如下三个 PMF 函数。

$mod_dur:修改时长。

$mod_bpit:修改韵律短语开始点的 f0。

$mod_ace:修改重音。

有两个变量的 PMF 函数设定的参数是绝对值。有三个变量的 PMF 函数设定的参数是相对默认值的相对值。PMF 函数中第一个变量表示该元素在词汇表中的位置。$mod_dur 的第三个变量表示相对默认时长变化的相对值。

③ 韵律操作符。当惯用模板中调用 PMF 函数时,在概念表述中给出韵律操作符。换句话说,PMF 与语句模式有关,而 IPS 和对话管理器根据上下文修改语音韵律参数和韵律操作符。SOCS 定义了如下三个韵律操作符。

$p_prominent(A):为了体现韵律上的重读,提高重音突出部分 50Hz,提高重音下降部分 40Hz,并把变量(A)的持续时间加大 5%。

$p_interrogative:提高疑问句句尾的 f0。SOCS 为其准备句尾的特定模式。

$p_speech(A,S):根据第二变量“S”控制第一变量 A 的发音速度。

图 6.14 表示图 6.12 中 s3 的韵律参数的生成结果。停顿标记 p2 为惯用模板 $why 在两个短语间插入一个长停顿。$mod_bpit 与 $why 一起将第二个韵律短语 kome(rice)开始点的 f0 设置为 190Hz。在 s3,$p_prominent 用于强调 samui(cold),并增加 samui 的 f0。

(5) 语音输出接口中的对话管理

图 6.11 中给出配合 SOCS 系统工作的语音输出接口。对话管理器能够修改 IPS 产生的概念表述。这样,输出语音就可与对话内容匹配、韵律自然。对话处理有如下两种。

① 抽取重读词语。按以下三种情况决定 IPS 中的词或短语是否重读:一是部分替代

或重复出现在表层句子中的词；二是 IPS 回答用户问题的要点；三是 IPS 使用的知识。第三种情况与对话管理无关，ISP 知道哪些词应该重读。

② 惯用模板的使用。在 SCOS 系统中，通常使用惯用模板。当使用 PMS 函数描述一个韵律模板时，在概念表述中使用惯用模板能合成更为自然的语音。ISP 不需要知道它能用的是什么惯用模板。对话管理器可以用惯用模板的形式改写 IPS 生成的概念表述。

## 6.7　听觉场景分析

听觉场景分析(Auditory Scene Analysis)源自契利(Cherry)在 1953 年的发现，即人的听觉系统能够从复杂的混合声音中有效地选择并跟踪某一说话人的声音(鸡尾酒会效应)，90 年代初由加拿大 McGill 大学著名听觉心理学家布莱曼(A S Bregman)提出的[100]。1995 年，德国波鸿鲁尔大学的马库斯(B Markus)提出了一种基于人的双耳听觉特性的双耳模型 CASA 方法[461]。把它用于噪声环境中语音识别的前端，实验结果表明采用了该前端处理后，语音识别的效果大为改善。1999 年，英国 Sheffield 大学的戈斯马克(D Godsmark)和布朗(J Brown)进一步发展了一种黑板模型的计算听觉场景分析模型[271]。自 20 世纪 90 年代以来国内学者开展了计算听觉场景分析的研究。

听觉场景分析是用来研究听觉系统如何对外界刺激进行组织与加工的。其任务有两个：一是找出那些能够使声谱成份组合到一起或使它们分离成独立的听觉流或表象的声学特征；二是研究听觉分组的方法。场景分析包含两个阶段，一是以格式塔原则为基础的初级分析，它把不同感觉元素分配到相应组中；另一阶段是图式加工，它可以对知觉组织进行验证和修复。这两个阶段分别对应于自下而上和自上而下两个处理过程。

### 6.7.1　初级分析

初级分析过程是先天的，无须有意注意参与。其策略是：先把听觉信号分割成许多独立的单元，这些单元与声谱中特定时域和频域相对应。然后，对这些单元进行分组或分离。分组是指听觉系统把某些具有相似特征或时间接近的音知觉为一个流，使之从复杂的环境声中突出出来。分离则是从复杂环境声中辨别出声音的不同来源或区分不同声音。分离和分组是一对统一的概念，如果出现了分组，也就意味着流与流之间产生了分离。初级分析包括序列整合和同时性整合。前者把在不同时间内顺序出现的谱成分纳入一个知觉流，以便计算环境中声音的序列特性。而后者则把同时出现的成分分开，将它们放入不同的流中。

#### 1. 序列整合

序列分组中的流形成遵循接近性和相似性原则，时间或频率接近的音将被分成一组。序列整合有两种形式，一种是对由两个音高不同的简单音交替出现构成的序列的整合，听者会把这个序列听成两个来源不同、分别由低音和高音构成的流。另一种形式是对由频率关系较为复杂的音所构成的序列进行整合。把两组频率关系变化的音，按一定顺序排列，构成一个序列，该序列能否产生曲调取决于音之间的频率关系。当两组音的频域相同时，曲调将消失；但如果它们分别在两个没有重叠的频率范围内，则曲调便被听成一个独立的流。影

响流分离最重要的因素有两个：一是交替速度；二是两个交替出现音之间的频率差。序列速度取决于音之间的时间间隔。布莱曼(A S Bregman)等[100]认为，产生分组效应的时间间隔为35ms。达尔温(C J Darwin)等[170]发现，当单个谐波提前或滞后32ms时，这个谐波就会与其他谐波产生分离。戴(H P Dai)等[169]甚至认为，听者可以按照特定任务要求，去调节时间整合间隔。但究竟是相同频率音之间的间隔重要，还是两个不同频率音之间的时间间隔重要，这个问题还有待进一步研究。

序列分组的影响因素还有基频、时间接近性、谱形状、强度和空间位置等。这些因素在分组中有竞争也有合作。如果所有因素对分组都有促进作用，则分组将被加强。例如，空间差异与其他因素联合起来时，其作用最强；又如，仅响度不同的两个音可能不会产生分离，但如果加入其他差异，则响度就可能起重要作用。

## 2. 同时性整合

声音中的谐波在频域中呈线性分布，而在基底膜上激活的相应位置则呈对数分布。在对数频率单元中，低次谐波之间相距较远，而高次谐波则相距较近。因此，谐波捕获有如下规律：(1)复合音中低次谐波比高次谐波更容易被捕获；(2)包含奇次谐波的谱成分比包含连续谐波的谱成分更容易被捕获；(3)相邻谐波被剔除的谐波易被捕获。因此谐波之间的频率相差越大，这些谐波越容易从复合音中被捕获。戴弗海斯(H Duifhuis)等[199]发现，在音高知觉中，听觉系统能够把额外声从复合音中剔除，其作用原理如同筛子，对信息进行过滤，他们把这种加工机制称作“谐波筛”(Harmonicsieve)。那些与基频很接近的谐波可以通过筛子；而其他额外声则无法通过。这种效应在低信噪比情况下更为突出。谐波筛是流形成的机制之一，其工作方式不是全或无(All-or-None)的，而是渐进的。在失调比例很小(3%～8%)的情况下，谐波仍然能够通过谐波筛。谐波筛只对那些可分辨的谐波起作用，它不能完全把额外音从元音的第一共振峰频域中剔除，因为两者之间可以产生部分整合。

调频和调幅对同时性整合同样产生影响。FM中，听觉系统使用FM范围差异对同时呈现的音进行知觉分离。FM有两种，一种是等差调制，它把原来的每个谐波都加上相同的频率。通过这种变化之后，它们之间的谐波关系消失，从而使音的不同成分产生分离。另一种是等比调制，即每个谐波被乘以相同的整数，这种处理之后，其谐波关系没有改变，但谐波之间的间距被扩大了。AM使不同谱位置上振幅产生变化。这种变化以及它们的出现时差和消失时差都对流分离产生影响。振幅的同步变化可以使谱产生分离，这与神经活动特征相一致。对应于不同谱位置的神经元的同步活动，保持时间很短。在频谱图上的每一段内，相对应的神经元同步活动，而段与段之间则不同步。音的识别就是通过对这些段中同步性变化的觉察而实现的。空间位置信息和频率信息的加工可能是独立的。生理学研究发现，猫的主要听觉皮层受损时，它对某个频率声音的空间位置判断能力将丧失，但仍能够判其他频率声音的空间位置。在人脑中也存在类似现象。另外，人可以同时听到不同位置上频率不同的两个纯音，两者不会融合。例如，在250～4000Hz内，最大频率差超过7%时，便不会产生融合。虽然听者把左右耳声音听成两个独立的音，但当双耳听到的音在频率上接近时，便出现双耳整合。

序列整合与同时性整合的差别是显而易见的。然而，两者都涉及声音的最基本特征：时域和频域，所以，它们不是孤立的。在复杂的声音知觉中，两者共同参与，相互影响。

## 6.7.2　以图式为基础的知觉组织

听者把环境中特定的声音信号，如言语、音乐以及其他熟悉的声音等存入记忆中，形成认知单元，这些认知单元就是认知图式。当听觉系统获得的信息模式与图式相同时，图式将被激活，并且通过图式对模式的其余部分进行推测。图式还可以被与其相关联的其他图式激活。图式加工是一个自上而下的加工过程，知识和注意在这一加工过程中起着重要作用。

### 1. 注意与知识的作用

图式加工需要对信息进行选择，它与注意关系密切。有意注意可以控制图式，只要任务要求注意参与，图式就会出现，即产生以图式为基础的分离。听觉系统可以利用频率线索把注意集中于一个特定的频率范围内。注意参与的加工过程可以很容易地把一个流分解，但不能对流之间的信息进行整合，因为注意只能指向一个流。图式形成过程就是获得关于刺激知识的过程。利用获得的知识，听者可以对刺激的变化趋势进行预测。当我们听一个重复的音时，就会掌握它的规则，形成图式。这些规则知识使我们心理上做好准备，把这一序列整合到连续的心理表征之中。由于规则可以很容易地使注意集中，所以它对流形态的影响很大。轨迹规则效应在有记忆参与的任务中最明显，以轨迹为基础的组织效果随刺激呈现次数增加而增加。听觉模式识别存在着图式成分的激活，这种激活受到时间规则和其他形式规则的影响。

### 2. 语音的加工

对言语声的组织同样有两种形式：序列整合和同时性整合。序列整合过程把按顺序出现的词(或元音)的各部分整合到一起，这是词识别的基础。产生这种整合的前提条件是相邻部分的声学特性(如：音高、共振峰以及基频等)具有连续性，或相差不大。语音的流分离与非语音的流分离一样，也受到序列速度的影响，速度越快分离程度也越大。另一种形式的组织是同时性整合，一般情况下，我们所面对的往往不是一个孤立的声音，而我们所能够意识到的可能只是其中的一个声音。这是因为听觉系统能够对众多的声音进行同时性整合，从而形成一个知觉流。同时性整合过程中，基频(或音高)是一个重要线索，差异越大，越容易分离。共振峰也表现出相同的作用规律。此外，声音的空间位置在同时性整合中也起一定的作用，不同位置的声音很容易被区分。与简单音模式相比，言语信号很不稳定，不仅它的基频随时间变化而变化，而且还包括与基频同步变化的口腔共振声。因此，语音加工和言语知觉过程更为复杂，加工过程中，既有初级分析，又存在着图式加工。一些初级分析模型只适用于简单元音或与元音相类似的音，而无法解释“鸡尾酒会效应”等复杂过程。如果用图式加工来解释则非常容易，即熟悉的声音或内容激活了记忆中相应的图式。

## 6.7.3　初级分析与图式加工之间的关系

初级分析和图式加工在功能上不是独立的，可能存在相互竞争。布莱曼等认为[100]：①图式加工在初级分析之后进行；②通过训练和学习，信息就可以用图式表征；③图式不参与听觉分析中的流形成过程；④图式描述的是知觉元素的典型特征，不论辅音、元音、音

节或其他言语成分都是如此。以格式塔原则为基础的听觉分组是低水平加工的结果。输入到听觉系统中的信息是杂乱的,这些无序波动将在听觉信息加工的后期阶段被最大限度地减弱。

两个过程对知觉的影响不同。初级分析过程把感觉信息分离,而图式加工过程则对信息进行选择,而不是将它从混合音中移走。初级分析中的分组是对称的,当它把高音和低音分离时,就形成两个独立的流。同样,也可以把两个来自不同空间位置的音分离开,判断出一个音来自左边,另一个来自右边。图式加工过程则没有这种对称性。在一个混杂环境中,我们可以很容易地听出自己的名字,但并不能辨认出在名字出现时,背景音是什么。因此,可以用分离的不对称性来确定是否出现了以图式为基础的分离。初级分析与图式加工的时间范围也不同,图式加工过程所涉及的时间范围比初级分析的时间范围大。由此可见,初级分析和图式加工是两个不同的阶段,前者是场景分析的第一阶段,当它不能解释言语组织过程时,就需要用图式来说明。把通过初级分析而产生的分离与图式加工结合起来,就可以避免或减少因初级分析和言语的熟悉性而产生的错误。两个系统所加工的感觉信息相同,但两者的加工难度不同,初级分析比图式加工更困难一些。

### 6.7.4 场景分析的总体评价

传统的听觉理论主要从生理学角度解释人的听觉过程,例如:地点学说、行波学说、齐射学说等,而场景分析从心理模型角度,把格式塔原则和图式过程应用于听觉信息加工。把听觉组织过程看成一个具有层次性的加工过程,丰富了听觉组织的理论。场景分析还用生态学的观点来分析听觉组织过程,从而增加了理论的外部效度,使之更加易于实际应用。然而,在某些理论与方法上,仍然存在一定局限性。

#### 1. 格式塔原则的局限性

当分组原则相冲突时,如果未产生融合,则格式塔原则是有效的。相反,如果所有的格式塔原则对分离或分组都有控制作用,并且产生知觉融合,那么非格式塔原则也必须参与,知觉组织才能顺利完成。由于每个人所发出的音都有其特定性,在言语知觉中,不是把那些简单的声音模式分组,而是要识别发音器官所发出的复杂声音;此外,还必须借助于已有的言语图式或其他知觉经验才能理解其含义。

#### 2. 关于图式的作用

布莱曼等只总结了那些有利于场景分析的研究结果,而忽略了一些反面的研究结果。格式塔理论在实验中很容易体现出来,而图式成分的作用则难以说明。场景分析把图式过程看成是万能的,不论何种条件下,一切格式塔原则所不能解释的现象都用它来解释。这种过分夸大图式作用的观点忽略了图式本身的重要特性。另外,图式的层次性、大小及相互关系等特征还有待于进一步研究。

图式的来源直接影响到对图式作用的认识。与初级分析相比,图式是学习获得的。乔姆斯基认为,要确认对儿童言语的强化是极端困难的,甚至是不可能的。所以,图式也并非完全都是后天习得的。

## 6.8 言语行为

英国哲学家奥斯汀(J L Austin)首先提出语言使用问题并进行认真的研究。他不仅首先探讨了语言使用问题，而且系统地、具体地研究了其中的一个问题，那就是说话本身为何就是一种行动[39]。奥斯汀认为，说任何一句话时，人们同时要完成三种行为：言内行为、言外行为、言后行为。言外行为是通过一定的话语形式，通过协定的步骤与协定的力而取得效果，所以言外行为是协定的。而言后行为依赖于语境，不一定通过话语本身就能取得，因此是不确定的。由于“言内行为”属于语言体系的范围，“言后行为”本身又不是语言行动，而且听者的反应也不是一个语言过程，而是复杂的心理过程，所以语言学家过去不大讨论“言后行为”，而把注意力集中在“言外行为”上。只是最近，一些学者才将语言问题与认知心理相结合，写出了颇具价值的文章，从而拓宽了语言的研究领域。奥斯汀把言外行为分为五类：即判定语(Verdictives)、裁定语(Exercitives)、承诺语(Commissives)、阐述语(Expositives)和行为语(Behabitives)。后来，Searle 修改了这一分类，把言外行为分为“新五类”：即，断言(Assertives)、指令(Directives)、承诺(Commissives)、表达(Expressives)和宣告(Declarations)[39]。

言语行为理论创立后立即引起了大家关注。其中美国哲学家塞尔(J R Searle)将言语系统化，阐述了言语行为的原则和分类标准，提出了间接言语行为这一特殊的言语行为类型。一个人直接通过话语形式的字面意义来实现其交际意图，这是直接的言语行为；当我们通过话语形式取得了话语本身之外的效果时，这就称作间接言语行为。简单地讲，间接言语行为就是通过做某一言外行为来做另一件言外行为。有四种基本的提出间接请求的方式：询问或陈述有关的能力、陈述一个愿望、陈述将来的行为以及列举出原因。

当和别人说话时，我们内隐地建立起了一种合作活动。事实上，如果我们交谈时彼此不默契，常常不知道彼此交谈的是什么，也没有交流到自己想说的。根据格利斯(H P Grice)的观点，成功的交谈遵守以下四个准则：数量准则、质量准则、相关准则和方式准则。

根据数量准则，你应该在谈话中按要求尽可能地提供信息，但也不能不恰当地提供过多信息。根据质量准则，你在对话中所说的应该是真实的。根据相关准则，你应该在谈话时注意所说的要和谈话的目的相关。根据方式准则，你应该表达清楚并避免生涩的表述、含糊的表达和有意混淆自己的观点。

言语行为理论的提出，无论对语言研究还是对应用语言学、社会语言学、语用学以及语言习的研究都产生了重大影响。一方面，它使学者们在有关方面的研究从以语法或语言形式为中心转向以言语功能为中心，从以单句为中心转向以语篇为中心，从以语言本身为中心转向以语言使用者、社团以及语言环境等为中心；另一方面，言语行为理论使诸多研究从以语言知识为中心转向以交际功能为中心。

# >>第 7 章
# 语言

语言是人类最重要的交际工具，是人们进行沟通交流的各种表达符号。语言是抽象思维的“物质外衣”，它是一种社会现象，劳动以及伴随劳动产生及发展起来的语言是产生人类思维的主要推动力量。

## 7.1 引言

语言是以语音为物质外壳、以词汇为建筑材料、以语法为结构规则而构成的体系。语言通常分为口语和文字两类。口语的表现形式为声音，文字的表现形式为形象。口语远较文字古老，人们学习语言也是先学口语，后学文字。口语的文法比较简单，所用词汇数量也比文字少。

语言是一种特殊的社会现象，它既不属于经济基础，也不属于上层建筑。它是人类在社会生产劳动过程中，由于交流思想的需要而产生的。语言与社会一直有着密切的联系。社会语言学就是研究语言结构变异与社会结构变化之间的相互关系。它的具体研究内容如下。

(1) 语言和社会。探索社会因素对语言的影响，以及语言因素对社会的影响。

(2) 语言变体。探索社会的变化如何引起语言在词汇和语法等方面的变化。

(3) 言语变体。从特定的言语环境中找出交际中言语的各种特点以及差异。

(4) 双语现象。从一个社会团体交替使用两种以上语言的情况中研究社会、心理等因素对语言选择的影响和制约。

(5) 话语分析。分析交际中的对话，探索组织谈话的社会、心理、语言、环境等结构及其内在联系，从而把语言研究扩大到言语领域之中。

(6) 语言功能。主要研究语言在社会中的各种功能，如交际功能、思维功能和传递信息的功能等等。

(7) 语言政策。根据社会发展的需要制定出科学的语言政策，进行语言建设。

思维同感觉和知觉一样，是人脑对客观现实的反映。不过，所反映的事物不是个别特征，而是一类事物的共同的、本质的特征。所谓间接的反映，就是说不是直接的，而是通过其他事物的媒介来反映客观事物。

语言和思维的关系问题是语言研究中最有趣、最富有争论性的问题之一，主要有下列观点。

行为主义心理学认为语言和思维是同一东西。华生认为思维与自言自语没有丝毫不同之处，他把思维完全看成是无声的语言。后来的新行为主义者斯金纳认为思维是无声的或

隐蔽的或微弱的言语行为。

另一种观点是语言决定思维。持这种观点的心理学家强调活动、语言、思维诸因素是相互联系向前发展的，也就是说，劳动及与其一起产生的语言，是思维、人类意识产生的最主要的推动力；各种活动、语言是个体思维产生的基础。这里所指的思维是语言思维，是以词为中介对现实的反映。语言思维是人类思维的基本形态。

20 世纪 30 年代，苏联著名生理学家巴甫洛夫(Иван Петрович Павлов)研究了神经系统的反射机制，提出了大脑中存在两种信号系统的学说。巴甫洛夫把人体在神经系统参与下对体外和体内刺激的反应即反射，区分为无条件反射和条件反射。在条件反射中，又把作用于人的各种条件刺激物分为第一信号系统和第二信号系统。凡是引起暂时神经联系的那些具体的条件刺激物叫第一信号；由第一信号在大脑皮质上所引起的暂时神经联系系统，叫第一信号系统。闻到菜的香味就会感到肚子饿，这就是第一信号系统活动。第一信号系统是人和动物所共有的。但人还有第二信号系统，这就是由语言文字这种信号在大脑皮质上所引起的暂时神经联系系统。如果吃到杨梅时感到酸，那么听别人讲到杨梅时也会引起酸的感觉，这就是第二信号系统活动。在这里，语言代替就是由于在人的大脑中存在上述两种信号系统及它们之间的相互作用。他说“如果我们对周围环境的感觉和表象对我们来讲是现实的第一信号，是具体的信号，那么言语，首先特别是从言语器官到大脑皮层去的运动刺激则是第二信号，即信号的信号。它们是现实的抽象，并且可以进行概括，这就组成了为人类所特有的高级思维，这种思维首先创立了一般人类的经验，而最后创立了科学，即人类对周围世界和对自己进行最高理解的工具”。

瑞士心理学家皮亚杰(Jean Piaget)从语言和思维的关系发生的起源史，从言语和思维的个体产生、发展趋势，从聋哑儿童与正常儿童的比较研究看到，逻辑思考的发生比语言、言语的发生要早，因此认为思维决定语言，语言是由逻辑所构成，逻辑运算从属于普通的动作协调规律，这些协调控制着所有的活动，包括语言本身在内。皮亚杰也承认语言在动作内化为表象和思维时起着主要作用，但也仅仅是影响内化的许多因素中的一种。

我们认为，思维和语言既是不同的概念，又是一个密切相关的统一体。思维是客观世界的主观映象，语言是对这种映象的表达形式。人类在劳动过程中由于协作的需要产生了语言，随着劳动的发展和社会的交往，人们的语言逐渐发展完善起来，思维也不断地发展和提高。思维的概括是借助于词，借助于语言来实现的。人们运用概念、判断、推理进行思维，反映事物的本质和规律性。有了概括化的语言，不仅可以把一代人联系起来，而且可以把不同世代的人联系起来，形成知识体系，使人的思维不仅是一种概括的反映，而且是一种以知识为中介的间接反映。

## 7.2　语言认知

人类借助语言交流思想，借助语言进行思维、推理。现已确定，语言功能定位于大脑左半球。与语言功能相关的大脑半球结构不对称在人类进化史上早在 30 万年前就已出现。人的语言潜能似乎在出生时就已存在。语言的普遍特征被认为部分起源于左半球语言相关皮质区域的特殊结构。从生物学意义上来讲，语言不是单一能力，而是一组能力，理解和表达就是其中的两个。语言理解和表达能力定位于左半球大脑皮质的不同区域。

语言学家和心理学家认为,语言习得中普遍性的机制是由人脑的结构决定的。根据这种观点,大脑的发育使人们具备了学习和运用语言的能力,而人们所讲的具体语种、方言和口音则是由社会环境决定的。图 7.1 给出了大脑语言信息处理的 Wernicke-Geschwind 模型[472]。

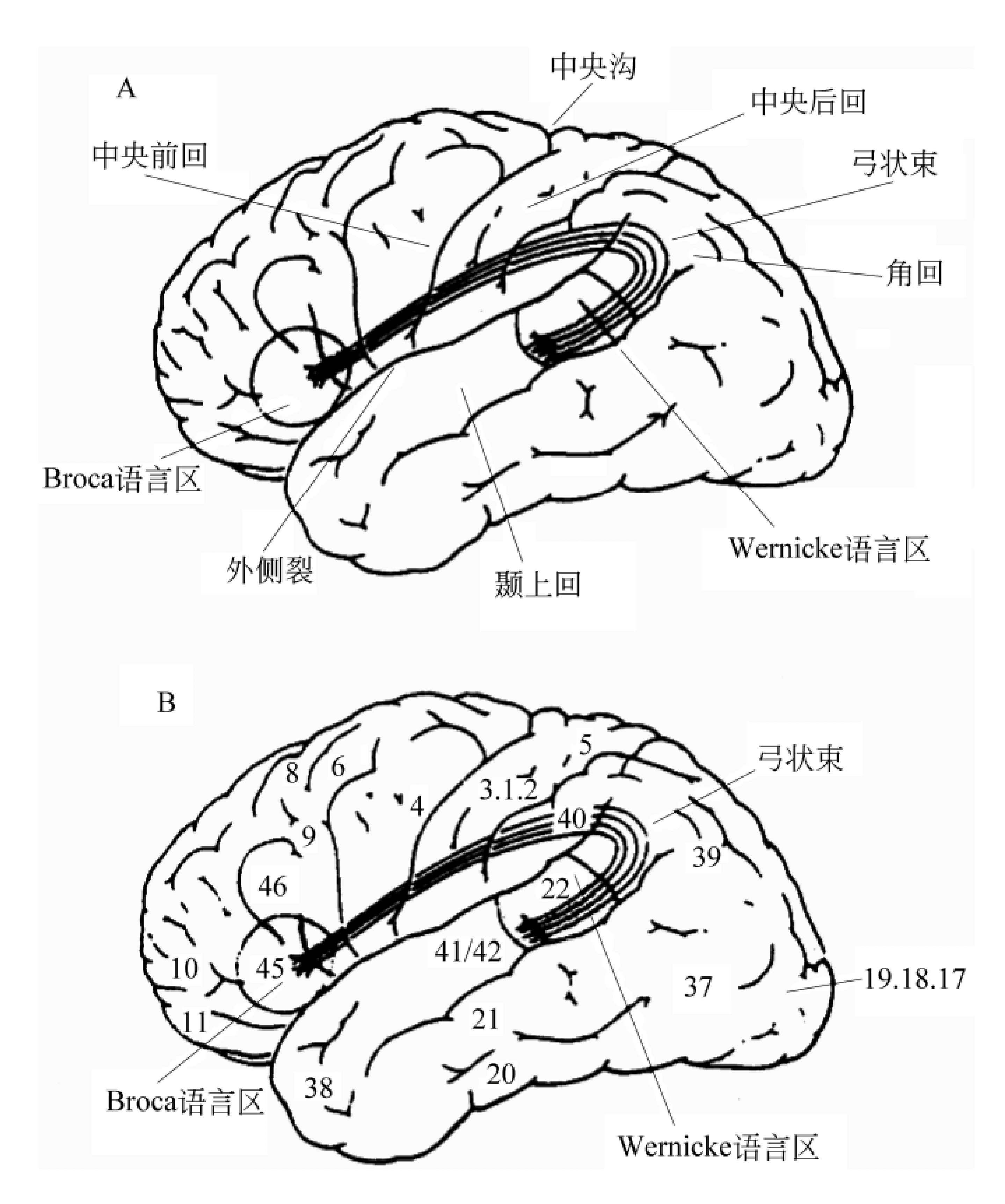



图 7.1 大脑语言信息处理的 Wernicke-Geschwind 模型

在 Wernicke-Geschwind 模型 A 中,给出左半球主要沟回及与语言功能相关的区域。Wernicke 语言区位于颞上回后部,靠近听觉皮层。Broca 语言区靠近运动皮层的面部代表区。连接 Wemicke 区和 Broca 区的通路称为弓状束。在 Wernicke-Geschwind 模型 B 中,给出左半球的 Brodmann 分区。41 区为初级听皮层,22 区为 Wernicke 语言区,45 区为 Broca 语言区,4 区为初级运动皮层。根据最初的 Wernicke-Geschwind 模型,人们听到一个词,信息自耳蜗基底膜经过听神经传至内侧膝状体,继而传至初级听皮层(Brodmann 41 区),然后至高级听皮层(42 区),再向角回(39 区)传递。角回是顶-颞-枕联合皮层的一个特定区域,被认为与传入的听觉、视觉和触觉信息的整合有关。由此,信息传至 Wernicke 区(22 区),进而又经弓状束传至 Broca 区(45 区)。在 Broca 区,语言的知觉被翻译为短语的语法结构,并储存着如何清晰地发出词的声音的记忆。然后,关于短语的声音模式的信息被传至控制发音的运动皮层面部表示区,从而使这个词能清晰地说出。

根据 Wernicke-Geschwind 模型可以作出临床上非常有用的预测。

(1) 预言了 Wernicke 区损伤导致的后果。到达听皮层的语言信息不能激活 Wernicke

区，因而将无法被理解。如果损伤向后方和下方扩展超越了 Wernicke 区，还将影响视觉性语言输入的处理。其结果是，患者对说的或写的语言都无法理解。

（2）该模型正确地预言了 Broca 区的损伤将不影响对说和写的语言的理解，但引起语言和词句生成的严重障碍，因为语言的声音模式和语言的结构模式不能传至运动皮层。

（3）该模型预言，弓状束损伤中断 Wernicke 区和 Broca 区的联系，将扰乱词语生成，因为听觉输入无法传至参与语言生成的脑区。

尽管 Wernicke-Geschwind 模型在临床上仍然有用，但是 Damasio，Raichle 和 Posner 等的认知和脑成像研究表明，Wernicke-Geschwind 模型过于简单化。语言功能涉及多个脑区以及这些脑区之间复杂的相互联系，并非由 Wernicke 区至 Broca 区及它们间的联系所能概括。语言障碍从来不像 Wernicke-Geschwind 模型所预测的那么单纯。图 7.2 是目前为止关于语言信息神经处理的一个较为理想的模型[472]。

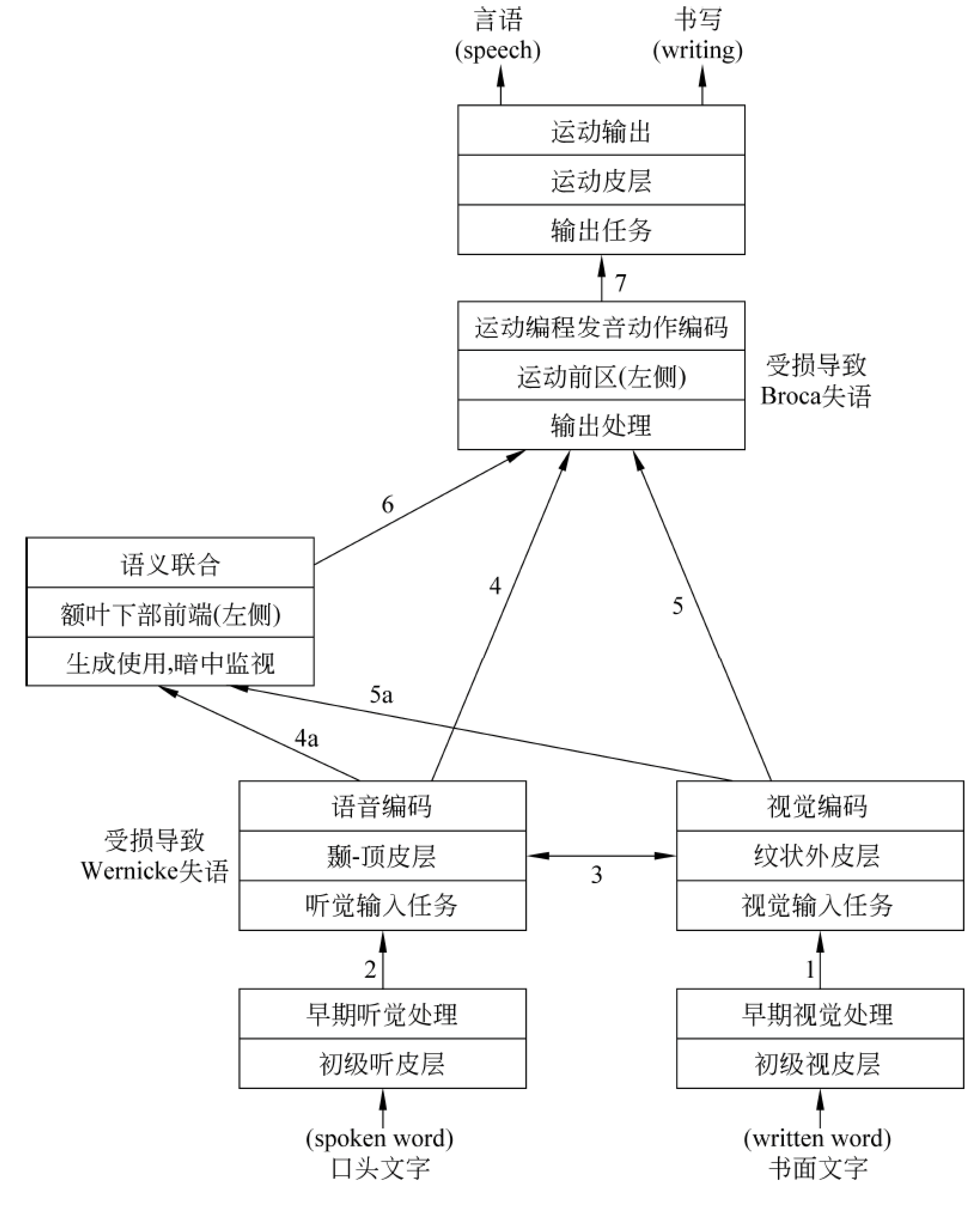



图 7.2　语言信息神经处理模型

根据 2004 年 8 月 1 日出版的美国《神经学》杂志介绍，日本以东京大学助理教授酒井邦义为首的科研小组发现大脑前皮层 Broca 区的一个特殊部分专门负责语法，另一区域则掌

管词汇的处理。科研人员在16名大学男生接受语法和词汇测验时,使用核磁共振成像技术来监视其脑部的活动。他们发现,Broca区临近左太阳穴的一个部分在语法判断时处于活跃状态,而更靠后的一处区域在记忆单词时受到了刺激。人脑负责语法的部分含有猴脑中不存在的组织,这可能有助于解释为什么只有人类才可以使用复杂的语言。这一发现将帮助人们深入了解失语症和痴呆症等疾病。

心理语言学是在心理学和语言学的结合点上产生的一门新兴学科,至今仅有近半个世纪的历史。其初期的研究主要受行为主义思想的支配,大多从刺激-反应的观点来探讨人类的语言行为,认为语言行为无非是一套习惯,通过刺激、反应和强化而逐渐形成。20世纪60年代,乔姆斯基提出了生成转换语法理论及表层结构和深层结构的概念。此后致力于证明乔姆斯基句法理论的心理现实性的研究,成为心理语言学研究中的主流。

到20世纪70年代,在继续研究句子的句法结构怎样认知加工的同时,开始注意研究句子的意义。研究问题也逐渐由单个句子发展到段落或课文的认知加工。“句子加工”和“课文加工”构成心理语言学两个最大、最主要的子领域。20世纪70年代和80年代,大多数关于课文加工的研究涉及某种推理在阅读时是否得到编码。例如:在阅读“青蛙坐在木板上,鱼在木板下面游。”之后问“鱼在青蛙下面游吗?”这样的推理性问题。但这里存在一个方法学问题,即不能区分推理是在阅读时进行的还是在推理测验时进行的。

福多尔(J A Fodor)提出了认知模块理论[226],与物理符号理论、人工神经网络、心理的生态学理论一样引起了人们的注意。认知模块被定义为一种快速的、强制的和封闭的信息过程。认知模块的最显著的特征是其信息过程的封闭性,即模块的活动与输出不受其他信息的影响。在认知模块理论看来,语言加工系统是由一系列在功能上彼此独立的模块组成的;每个模块是独立的加工单位;加工是自动的、强制的;这个过程不受其他模块的影响。例如,词汇加工或句法加工作为独立的模块,其加工不受句子语境等高层次因素的影响。

鸠斯巴切(M A Gernsbacher)提出的结构构建框架是一个较著名的语言理解模型[256],该理论认为,理解的目的是构造一个连贯的心理表征或结构,表征的建构材料是记忆细胞。这一建造过程包括三个步骤:奠基,即根据最初输入的信息形成一个基础结构;映射,即当新输入的信息与先前信息一致时,就映射到这个基础上去,从而发展原来的结构;转移,即当新输入的信息与先前信息不一致时,就开始构建一个新的子结构。因此,大多数心理结构都是由若干分支的子结构组成的。在结构建造过程中,有两种机制控制记忆细胞的激活水平:增强,即与正在建构的结构一致的信息的激活水平的提高;压抑,即与正在建构的结构不一致的信息的激活水平的主动下降。该理论强调,压抑不适当信息和干扰信息是有效理解的基础,压抑机制的效率可能是语言理解能力中个别差异的重要原因。

上面仅讨论了语言的认知成分。然而,人类的交流还具有重要的情感成分。这些成分包括音乐性的语律和情感性的姿态。语言的某些情感成分依赖于右半球的专门处理。与右半球损伤相关的语言情感成分丧失称为语韵缺失。语韵处理在右半球的组织方式与语言的认知方面在左半球的组织方式似乎相对应。右半球前部损伤的患者无论是悲或是喜,讲话时语调总是平板的。右半球后部损伤的患者则不能理解别人语言中的情感成分。

# 7.3　乔姆斯基的形式文法

在计算机科学中，形式语言是某个字母表上一些有限长字串的集合，而形式文法是描述这个集合的一种方法。形式文法之所以这样命名，是因为它与人类自然语言中的文法相似。最常见的文法的分类系统是乔姆斯基于 1950 年发展的乔姆斯基谱系，这个分类谱系把所有的文法分成四种类型：短语结构文法、上下文有关文法、上下文无关文法和正规文法。任何语言都可以由无限制文法来表达，余下的三类文法对应的语言类分别是递归可枚举语言、上下文无关语言和正规语言[138]。依照排列次序，这四种文法类型依次拥有越来越严格的产生式规则，所能表达的语言也越来越少。尽管表达能力比短语结构文法和上下文相关文法要弱，但由于能高效率的实现，上下文无关文法和正规文法成为四类文法中最重要的两种文法类型。

## 7.3.1　短语结构文法

短语结构文法是一种非受限文法，也称为 0 型文法，是形式语言理论中的一种重要文法。一个四元组 $G=(\Sigma,V,S,P)$，其中 $\Sigma$ 是终结符的有限字母表，$V$ 是非终结符的有限字母表，$S(\in V)$是开始符号，$P$ 是生成式的有限非空集，$P$ 中的生成式都为 $\alpha\to\beta$ 的形式，这里 $\alpha\in(\Sigma\cup V)^*V(\Sigma\cup V)^*$，$\beta\in(\Sigma\cup V)^*$。短语结构文法又称为 0 型文法。因对 $\alpha$ 和 $\beta$ 不加任何限制，故也称其为无限制文法。0 型文法生成的语言类与图灵机接受的语言类相同，称为 0 型语言类（常用 $L_0$ 表示）或递归可枚举语言类（常用 $Lre$ 表示）。

**表 7.1　$L_0$，$L_1$ 在代数运算下的封闭性**

| 代数运算 \ 语言类 | $L_0$ | $L_1$ |
|---|---|---|
| 求并 | √ | √ |
| 求连结 | √ | √ |
| 求闭包 | √ | √ |
| 求补 | × | ? |
| 求交 | √ | √ |
| 与正则语言相交 | √ | √ |
| 反演 | √ | √ |
| 置换 | √ | × |

注：√表示封闭，×表示不封闭，? 表示尚未解决。

例如：$G=(\{a\},\{[,],A,D,S\},S,P)$，其中：$P=\{S\to[A],[\to[D,D]\to],DA\to AAD,[\to\wedge,]\to\wedge,A\to a\}$，显然，$G$ 是短语结构文法，它所生成的语言 $L(G)=\{a^{2^n}\mid n\geqslant 0\}$ 是 0 型语言。

0 型语言在一些代数运算下的封闭性如表 7.1 所示，关于判定问题的一些结果如表 7.2 所示。表中 $D$ 表示可判定，$U$ 表示不可判定，$G$ 表示文法，$L$ 表示语言。

表 7.2　$L_0$,$L_1$ 有关的判定问题

| 判定问题 \ 语言类 | $L_0$ | $L_1$ |
|---|---|---|
| 任意字 $x\in L(G)$? | $U$ | $D$ |
| $L(G_1)\subset L(G_2)$? | $U$ | $U$ |
| $L(G_1)=L(G_2)$? | $U$ | $U$ |
| $L(G)=\varnothing$? | $U$ | $U$ |
| $L(G)=$无限集合? | $U$ | $U$ |
| $L(G)=\Sigma^*$ ? | $U$ | $U$ |

短语结构文法的标准型为:$A\to\xi$,$A\to BC$,$A\to\wedge$,$AB\to CD$,其中 $\xi\in(\Sigma\cup V)$,$A,B,C,D\in V$,$\wedge$ 是空字。

对短语结构文法中的生成式作某些限制,即得到上下文有关文法、上下文无关文法和正则文法。

## 7.3.2　上下文有关文法

上下文有关文法是形式语言理论中的一种重要文法。一个四元组 $G=(\Sigma,V,S,P)$,其中 $\Sigma$ 是终结符的有限字母表,$V$ 是非终结符的有限字母表,$S(\in V)$是开始符号,$P$ 是生成式的有限非空集,$P$ 中的生成式都为 $\alpha A\beta\to\alpha\gamma\beta$ 的形式,这里 $A\in V$,$\alpha,\beta\in(\Sigma\cup V)^*$,$\gamma\in(\Sigma\cup V)^+$。上下文有关文法又称为1型文法。其生成式的直观意义是:在左有 $\alpha$,右有 $\beta$ 的上下文中,$A$ 可以被 $\gamma$ 所替换。上下文有关文法所生成的语言称为上下文有关语言或1型语言。常用 $L_1$ 表示1型语言类。

单调文法,若文法 $G=(\Sigma,V,S,P)$的所有生成式都为 $\alpha\to\beta$ 的形式并且 $|\alpha|\leqslant|\beta|$,其中 $\alpha\in(\Sigma\cup V)^*V(\Sigma\cup V)^*$,$\beta\in(\Sigma\cup V)^+$,则称 $G$ 为单调文法。单调文法可简化使 $P$ 中任意生成式的右侧长最大为2,即:若 $\alpha\to\beta\in P$,则 $|\beta|\leqslant 2$。已经证明:单调文法所生成的语言类与1型语言类,即上下文有关语言类相同。因此,有的文献把单调文法的定义作为上下文有关文法的定义。

例如:$G=(\{a,b,c\},\{S,A,B\},S,P)$,其中:$P=\{S\to aSAB/aAB,BA\to AB,aA\to ab,bA\to bb,bB\to bc,CB\to cc\}$,显然,$G$ 是单调文法,因而也是上下文有关文法。它所生成的语言 $L(G)=\{a^nb^nc^n\mid n\geqslant 1\}$是上下文有关语言。

上下文有关文法的标准型为:$A\to\xi$,$A\to BC$,$AB\to CD$,其中 $\xi\in(\Sigma\cup V)$,$A,B,C,D\in V$。上下文有关语言类与线性有界自动机接受的语言类相同。1型语言对运算的封闭性以及关于判定问题的一些结果参见短语结构文法中的表1和表2。特别要指出的是:1型语言对补运算是否封闭是迄今未解决的一个问题。

## 7.3.3　上下文无关文法

上下文无关文法是形式语言理论中一种重要的变换文法,在乔姆斯基分层中称为2型文法,生成的语言称为上下文无关语言或2型语言,在程序设计语言的语法描述中有重要应用。

上下文无关文法(简称 CFG)可以化为两种简单的范式之一,即任一上下文无关语言(简称 CFL)可用如下两种标准 CFG 的任意一种生成:其一是乔姆斯基范式,它的产生式均取 $A\rightarrow BC$ 或 $A\rightarrow a$ 的形式;其二是格雷巴赫范式,它的产生式均取 $A\rightarrow aBC$ 或 $A\rightarrow \alpha$ 的形式。其中 $A,B,C\in V$,是非终结符;$a\in\Sigma$,是终结符;$\alpha\in\Sigma^*$,是终结符串。

从文法生成语言,可有多种推导方式。例如文法$\{S\rightarrow AB, A\rightarrow a, B\rightarrow b\}$可有两种推导:$S\Rightarrow AB\Rightarrow aB\Rightarrow ab$ 及 $S\Rightarrow AB\Rightarrow Ab\Rightarrow ab$。若每次都取最左边的非终结符进行推导,如上例中的前一种方式那样,则称为左推导。如果有两种不同的左推导推出同一结果,则称此文法是歧义的,反之是无歧义文法。对有些歧义文法,可找到一个等价的无歧义文法,生成同一个语言。不具有无歧义文法的语言称为本质歧义性语言。例如,$\{S\rightarrow A,\ S\rightarrow a, A\rightarrow a\}$是歧义的文法。$L=\{a^m b^n c^n \mid m,n\geqslant 1\}\cup\{a^m b^m c^n \mid m,n\geqslant 1\}$是本质歧义性语言。接受 CFL 的自动机称为下推自动机。确定型和不确定型下推自动机接受的语言分别称为确定型 CFL 和不确定型 CFL,前者是后者的真子集。例如,$L=\{a^n b^n \mid n\geqslant 1\}\cup\{a^n b^{2n} \mid n\geqslant 1\}$是一个不确定型 CFL 而不是确定型 CFL。

对任意正整数 $n$,令 $\Sigma_n=\{a_1,\cdots,a_n\}$,$\Sigma_n'=\{a_1',\cdots,a_n'\}$,定义乔姆斯基变换文法 $G=(\Sigma,V,S,P)$为$(\Sigma_n\cup\Sigma_n',\{S\},S,\{S\rightarrow,Sa_iSa_i'S\mid 1\leqslant i\leqslant n\})$。这个文法生成的语言称为代克集。如果把 $a_i$ 看作开括号。把 $a_i{}'$看作相应的闭括号,则 $n$ 维代克集 $D_n$ 就是由 $n$ 种不同的括号对组成的配对序列之集合。例如,$a_1a_2a_2a_2'a_2'a_1'$和 $a_1a_1'a_2a_2'a_1a_1'$都属于 $D_2$。

代克集是把正则语言族扩大成上下文无关语言族的工具。对任一上下文无关语言 $L$,必存在两个同态映射 $h_1$ 和 $h_2$,以及一个正则语言 $R$,使 $L=h_2[h_1^{-1}(D_2)\cap R]$,其中 $D_2$ 是二维代克集,反之亦然。

更进一步,上下文无关语言族是包含 $D_2$,且在同态、逆同态和与正则语言相交三种代数运算下封闭的最小语言族。

由于上下文无关文法被广泛地应用于描述程序设计语言的语法,因此更重要的是从机械执行语法分解的角度取上下文无关文法的子文法,最重要的一类就是无歧义的上下文无关文法,因为无歧义性对于计算机语言的语法分解至为重要。在无歧义的上下文无关文法中最重要的子类是 $LR(k)$文法,它只要求向前看 $k$ 个符号即能作正确的自左至右语法分解。$LR(k)$文法能描述所有的确定型上下文无关语言,但是对于任意的 $k>1$,由 $LR(k)$文法生成的语言必可由一等价的 $LR(1)$文法生成。$LR(0)$文法生成的语言类是 $LR(1)$文法生成的语言类的真子类。

## 7.3.4　正则文法

正则文法来源于 20 世纪 50 年代中期乔姆斯基对自然语言的研究,是乔姆斯基短语结构文法分层里的 3 型文法。正则文法类是上下文无关(2 型)文法类的真子类,已应用于计算机程序语言编译器的设计、词法分析(文本处理中描述触发过程动作的文本模式、文件类型和扫描器、文本工具的标准基础)、开关电路设计、句法模式识别等,是计算机和信息科学、工程、物理、化学、生物、医学、应用数学不可忽视的论题。

正规文法有多种等价的定义,我们可以用"左线性文法"或者"右线性文法"来等价地定义正规文法。"左线性文法"要求产生式的左侧只能包含一个非终结符号,产生式的右侧只能是空串、一个终结符号或者一个非终结符号后随一个终结符号。"右线性文法"要求产生

式的左侧只能包含一个非终结符号,产生式的右侧只能是空串、一个终结符号或者一个终结符号后随一个非终结符号。

一个左线性文法可用四元组 $G=(V,\Sigma,P,S)$ 表示,其中 $V$ 是变元的有限集合,$\Sigma$ 是终结符的有限集合,$S\in V$,称为开始符号,$w\in\Sigma^*$(即 $w$ 为有限个终结符连接成的串或字,可能为空串或空字 $\varepsilon$)。$A,B\in V$ 时,$P$ 是由形为 $A\to w$ 和 $A\to wB(A\to Bw)$ 产生式组成的有限集。右线性文法与左线性文法是等价的,即可生成同样的语言(字集合)类。

正则文法的结构与复杂性测度由变元、产生式的个数及文法有向图的高度、每一层的结点数来确定。$S|\frac{*}{G}w$ 表示有限次使用 $P$ 中产生式可派生出字 $w$,正则文法 $G$ 可作为生成器产生和描述正则语言 $L(G)=\left\{w\in\Sigma^* \mid S|\frac{*}{G}w\right\}$。例如 $G=(\{S,A,B\},\{0,1\}P,S)$,$P=\{S\to 0A|0,A\to 1B,B\to 0A|0\}$,$G$ 是一个正则(右线性)文法,$L(G)$ 中含字 0,$(S\to 0)$,01010$(S\to 0A\to 01B\to 010A\to 0101B\to 01010)$。正则语言也称为正则集,可以用正则表达式表示。对任一正则表达式,可以构造出带 $\varepsilon$ 动作的非确定有限自动机(NFA)在线性时间内来接受它,也可构造出不带 $\varepsilon$ 动作的确定有限自动机(DFA)在平方时间内来接受它,正则文法生成的语言也可由双向确定有限自动机(2DFA)来接受,NFA、DFA、2DFA 是等价的,即所接受的语言类是相同的。

正则表达式递归地定义为,设 $\Sigma$ 为有限集,

(1) $\varnothing$,$\varepsilon$ 和 $a(\forall a\in\Sigma)$ 是 $\Sigma$ 上的正则表达式,它们分别表示空集、空字集 $\{\varepsilon\}$ 和集合 $\{a\}$。

(2) 若 $\alpha$ 和 $\beta$ 是 $\Sigma$ 上的正则表达式,则 $\alpha\cup\beta$,$\alpha\cdot\beta=\alpha\beta$ 和 $\alpha^*$ 也是 $\Sigma$ 上的正则表达式,它们分别表示字集 $\{\alpha\}$,$\{\beta\}$,$\{\alpha\}\cup\{\beta\}$,$\{\alpha\}\{\beta\}$ 和 $\{\alpha\}^*$,(运算符 $\cup$,$\cdot$,$*$ 分别表示并、连接和星$\left(\text{乘幂闭包}\{\alpha\}^*=\left\{\bigcup_{i=0}^{\infty}\alpha^i\right\}\right)$,优先顺序为 $*$,$\cdot$,$\cup$。

(3) 只有有限次使用(1)和(2)确定的表达式才是 $\Sigma$ 上的正则表达式,只有 $\Sigma$ 上的正则表达式所表示的字集才是 $\Sigma$ 上的正则集。

正则文法生成的字集,其正则表达式为 0。为了简化正则表达式,常用下列等式。

(1) $\alpha\cup\alpha=\alpha$(幂等律)。

(2) $\alpha\cup\beta=\beta\cup\alpha$(交换律)。

(3) $(\alpha\cup\beta)\cup\gamma=\alpha\cup(\beta\cup\gamma)$(结合律)。

(4) $\alpha\cup\varnothing=\alpha$,$\alpha\varnothing=\varnothing\alpha=\varnothing$,$\alpha\varepsilon=\varepsilon\alpha=\alpha$(零一律)。

(5) $(\alpha\beta)\gamma=\alpha(\beta\gamma)$(结合律)。

(6) $(\alpha\cup\beta)\gamma=\alpha\gamma\cup\beta\gamma$(分配律)。

(7) $\varepsilon\cup\alpha^*=\alpha^*$。

(8) $(\varepsilon\cup\alpha)^*=\alpha^*$。

用(1)至(2)可将 $\alpha$ 变为 $\beta$ 时,称 $\alpha$ 与 $\beta$ 相似。

运用正则表达式方程 $X_i=a_{i0}+a_{i1}X_1+\cdots+a_{in}X_n$ 来处理语言有其方便之处,因为这种方程中 $\Delta=\{X_1,\cdots,X_n\}$(未知量的集合)与 $\Sigma$ 之交为 $\varnothing$,$\alpha_{ij}$ 为 $\Sigma$ 上正则表达式,当 $\alpha_{ij}$ 为 $\varnothing$,$\varepsilon$ 时分别相当于普通线性方程组之系数 0、1,可以按线性方程组的高斯消去法求解,当然,这里的解是

一个集合,即解不是唯一的,但该算法能够正确地确定一个极小不动点作解。

由所生成的正则语言来体现。如果 $R$ 为正则语言,则存在一个常数 $n$,使得 $R$ 中所有字长不小于 $n$ 的字 $w$ 都可写成 $xyz$ 的形式($y \neq \varepsilon$ 且 $|xy| \leqslant n$)并且对所有非负整数 $i$ 必有 $xy^iz \in R$,此为泵引理。它是证明某些语言不正则的有力工具,且有助于建立算法来判断一个给定的正则文法所生成的语言是有限的还是无限的。判断某些语言是否正则还可以利用对语言运算是否封闭来决定。已知正则语言类对布尔运算(并、交、补)、连接、*(克林尼闭包)、左右商、替换、同态、逆同态、INIT(求前缀)、FIN(求后缀)、MIN、MAX、CYCLE、Reversal 等封闭。又当 $p(x)$ 为非负整系数多项式,$R$ 为正则语言时,$L_1 = \{w \mid$ 对某个使 $|y| = p(|w|)$ 的 $y$ 有 $wy \in R\}$,$L_2 = \{w \mid$ 对某个使 $|y| = p(|w|)$ 的 $y$ 有 $wy \in R\}$ 也是正则语言。当 $R$,$R_1$ 和 $R_2$ 是正则语言时下列问题都是可判定的:$w \in R$? $R = \varnothing$ ? $R = \Sigma^*$ ? $R_1 = R_2$ ? $R_1 \subseteq R_2$ ? $R_1 \cap R_2 = \varnothing$ ?

## 7.4 扩充转移网络

1970 年,美国人工智能专家伍兹(W Woods)研究了一种语言自动分析的方法,叫做扩充转移网络(Augmented Transition Networks,ATN)[818]。ATN 是在有限状态文法的基础上,作了重要的扩充之后研制出来的。有限状态文法可以用状态图来表示,但这种文法的功能仅在于生成。如果从分析句子的角度出发,我们也可以用状态图来形象地表示一个句子的分析过程,这样的状态图叫做有限状态转移图(FSTD)。一个 FSTD 由许多个有限的状态以及从一个状态到另一个状态的弧所组成,在弧上只能标以终极符号(即具体的词)和词类符号(如< Verb >、< Adj >、< Noun >等,分析从开始状态出发,按着有限状态转移图中箭头所指的方向,一个状态一个状态地扫描输入词,看所输入的词与弧上的标号是否相配,如果扫描到输入句子的终点,FSTD 进入最后状态,那么,FSTD 接收了输入句子,分析也就完成了(见图 7.3)。

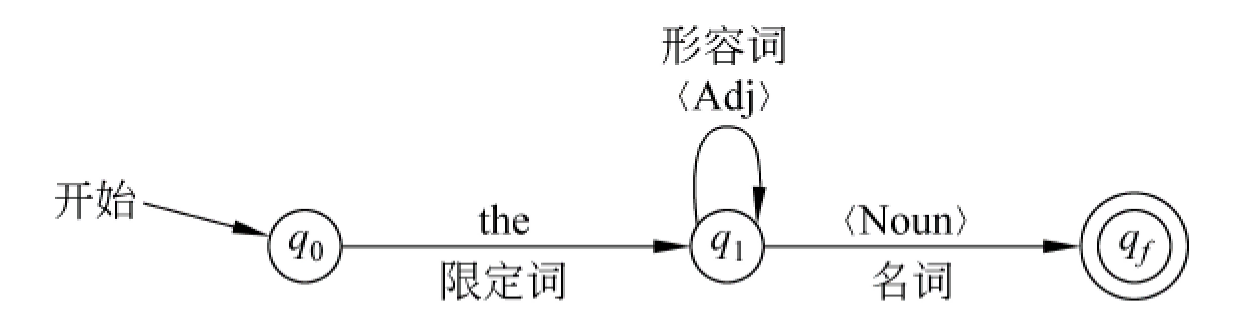



图 7.3 扩充转移网络转移图

扩充转移网络只能识别有限状态语言。我们知道,有限状态文法的重写规则是 A→aQ 或 A→a,这种文法是比较简单的,FSTD 有足够的能力来识别由有限状态文法生成的语言。

例如,我们可以提出这样一个由 FSTD 来分析的名词词组,其中以 the 开头,以< Noun >结尾,开头和结尾可有任意个< Adj >。如:

the pretty picture(美丽的图画)
the old man(老人)
the good idea(好主意)

FSTD 如图 7.3 所示。如果输入的名词词组是 the pretty piculre,从状态 $q_0$ 开始,沿着标有 the 的弧进行扫描,因为 the 是输入符号串的最左词,二者相匹配,然后进入状态

$q_1$,而在输入符号串中剩下来应该分析的部分是 pretty picture,在走过标有< Adj >的这个成圈的弧后,我们又进入状态 $q_1$,而在输入符号串中剩下的部分是 picture,由于该词是名词,与弧上的标号< Noun >相配,故进入最后状态 $q_f$。这时,输入符号串中的全部的词都检查完毕,分析结果,FSTD 接收了这个符号串。

有限状态文法是不适于处理很复杂的自然语言的,因此,有必要对 FSTD 加以扩充,给它提供一个递归的机制,增加其识别能力,以便处理上下文自由语言。为此,提出递归转移网络(Recursive Transition Networks,RTN)。RTN 也是一个有限状态转移图,不过,其中弧上的标记不仅可以包含终极符号(即具体的词)和词类符号,而且,还可以是词组类型符号(如 NP、S、PP 等)。由于每个词组类型符号又可以用另一个有限状态转移图来表示,这样,RTN 便具有递归的能力,每当扫描到词组类型时,RTN 可临时转移到与该词组类型相应的另一个有限状态转移图中,以便对分析过程进行临时控制。这样,RTN 不仅能识别有限状态语言,而且,还能识别上下文无关语言,扩大了 FSTD 的识别能力(见图 7.4)。

RTN 的操作方式与 FSTD 的操作方式类似。如果弧上的标记是终极符号或词类符号,那么,可以像 FSTD 那样处理该弧。例如 ball 这个词,可与标记为< Noun >的弧相匹配,而不能与标记为< Adj >的弧相匹配。如果弧上的标记是词组类型符号,而这个词组类型符号又与另一个有限状态转移图相对应,那么,就把当前的分析状态置入栈中,控制转移到相应名字的有限状态转移图中,继续处理这个句子。当处理结束或处理失败时,控制又转回来,回到原来的那个状态继续进行处理。

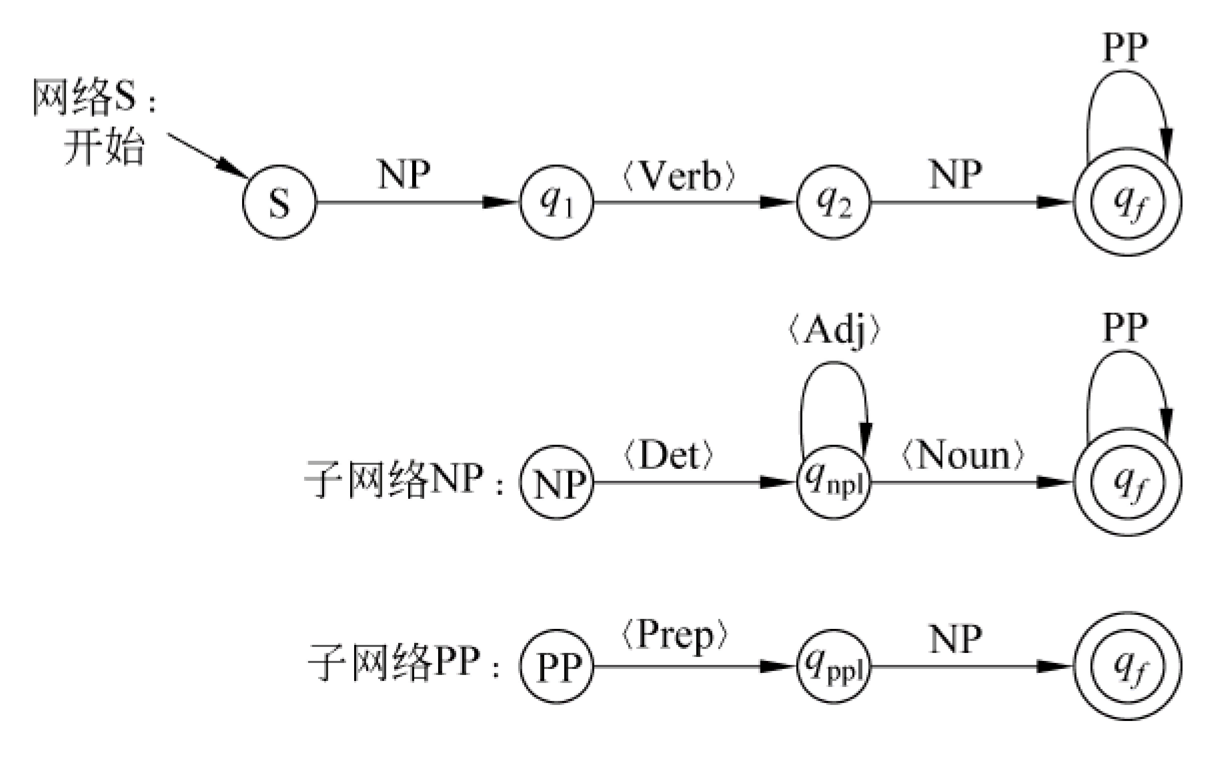



图 7.4　RTN 网络

例如,设 RTN 由名字为 S 的网络,以及名字为 NP 及 PP 的两个子网络构成。这里,NP 表示名词词组,PP 表示前置词词组,< Det >是限定词,< Prep >是前置词,< Adj >是形容词,< Noun >是名词,最后状态标以 $q_f$。如果输入符号串是 the little boy in the swimsuit kicked the red ball (那个穿游泳装的小男孩踢了那个红色的球),上面的 RTN 将按如下的顺序来进行分析:

```
NP: the little boy in the swimsuit
PP: in the Swimsuit
NP: the Swimsuit
Verb: kicked
NP: the red ball
```

在网络 S 中,从 S 出发,扫描到 NP,于是控制进入名字为 NP 的子网络处理 the little

boy in the Swimsuit 这个 NP，当扫描完 the 1ittle boy 后，在子网络中还可扫描到 PP，即 in the swimsuit。于是，控制进入名字为 PP 的子网络处理 in the swimsuit 这个 PP 中，在这个子网络中，当扫描完< Prep >即 in 后，就应扫描 NP the swimsuit，于是，控制又进入名字为 NP 的子网络，处理 the swimsuit 这个 NP，进入这个名字为 NP 的子网络的最后状态，于是，名词词组 the little boy in the swimsuit 处理完毕，控制回到网络 S，进入状态 $q_1$，在状态 $q_1$ 扫描动词 kicked，进入状态 $q_2$，在状态 $q_2$，又扫描名词词组 NP，于是，控制又进入名字为 NP 的子网络处理名词词组，the red ball，处理完这个名词词组，才进入网络 S 中的最后状态 $q_f$，句子分析完毕。

RTN 能处理上下文无关语言。但是，我们知道，能生成上下文无关语言的上下文无关文法对于处理自然语言仍然是不完善的。因此，还得进一步扩充 RTN，使之具有更强的识别能力。这样，Woods 便提出了扩充转移网络，即 ATN。ATN 是由 RTN 作了如下三方面的扩充而成的。

(1) 加一个寄存器，以便储存信息。例如，在不同的子网络之间，可能会局部地形成一些推导树，这样的推导树便可暂时储存于寄存器中。

(2) 网络中的弧除了可以标记终极符号、词类符号、词组类型符号之外，还可以检查是否满足进入这个弧的条件。

(3) 在弧上还可以执行某些动作，重新安排句子的结构。

由于加上了寄存器、条件及动作，ATN 的功能可提高到图灵机的水平。从理论上说，ATN 有足够的能力来识别可以用计算机识别的任何语言。

ATN 的操作方式与 RTN 类似。不同之处在于：如果在弧上标有“检查”，就得首先执行这个“检查”，仅当“检查”成功时，才可继续扫描该弧，另外，如果弧上要执行与之相关的动作，那么，在扫描完这个弧后，就得再执行这些动作。目前，ATN 已被成功地应用于人机对话的研究中，也可以用于文句生成。

ATN 也有一些局限性，它过分地依赖于句法分析，限制了它处理某些含语义但不完全合语法的话语的能力。

## 7.5 格文法

格文法(Case Grammar)是由费尔蒙(C J Fillmore)提出的[224]，主要是为了找出动词和跟它处在结构关系中的名词的语义关系，同时也扩及动词或动词短语与其他的各种名词短语之间的关系。也就是说，格文法的特点是允许以动词为中心构造分析结果，尽管文法规则只描述句法，但分析结果产生的结构却相应于语义关系，而非严格的句法关系。例如，对于英语句子：

```
Mary hit Bill
```

的格文法分析结果可以表示为：

```
(hit  ( Agent Mary )
      ( Dative Bill ))
```

这种表示结构称为格文法。在格表示中,一个语句包含的名词词组和介词词组均以它们与句子中动词的关系来表示,称为格。上面的例子中 Agent 和 Dative 都是格,而像“(Agent Mary)”这样的基本表示称为格结构。

在传统语法中,格仅表示一个词或短语在句子中的功能,如主格、宾格等,反映的也只是词尾的变化规则,故称为表层格。在格文法中,格表示的语义方面的关系,反映的是句子中包含的思想、观念等,称为深层格。和短语结构语法相比,格文法对于句子的深层语义有着更好的描述。无论句子的表层形式如何变化,如主动语态变为被动语态,陈述句变为疑问句,肯定句变为否定句等,其底层的语义关系,各名词成分所代表的格关系不会发生相应的变化。例如,被动句“Bill was hit by Mary”与上述主动句具有不同的句法分析树(如图 7.5 所示),但格表示完全相同,以说明这两个句子的语义相同,并实现多对一的源-目的映射。

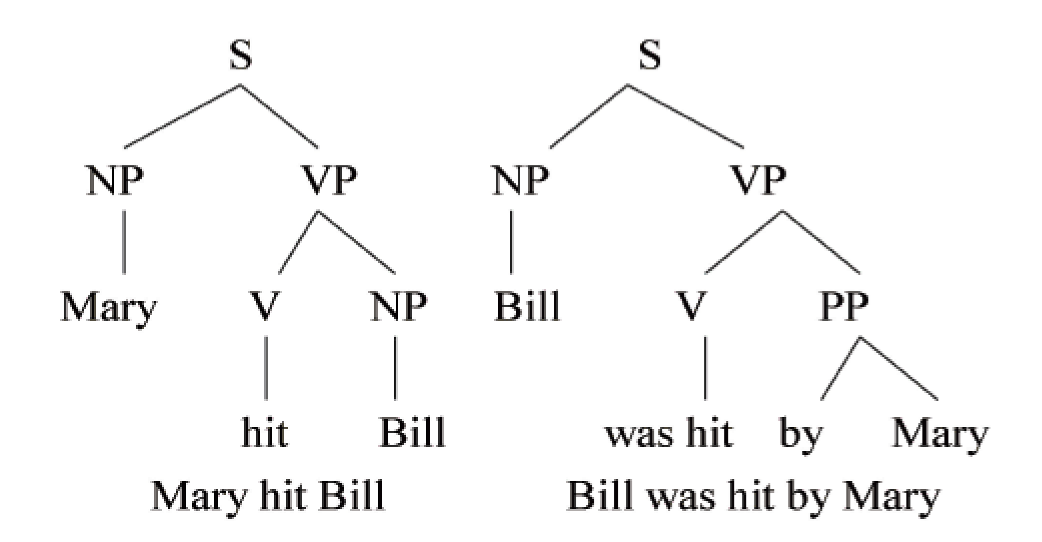



图 7.5　主动句和被动句的句法分析树

格文法和类型层次相结合,可以从语义上对 ATN 进行解释。类型层次描述了层次中父子之间的子集关系,或者说,父结点比子结点更一般。根据层次中事件或项的特化(Specialized)/泛化(Generalized)关系,类型层次在构造有关动词及其宾语的知识,或者确定一个名词或动词的意义时非常有用。

在类型层次中,为了解释 ATN 的意义,动词具有关键的作用。因此可以使用格文法,通过动作实施的工具或手段(Instrument)来描述动作主体(Agent)的动作。例如,动词 laugh 可以是通过动作主体的嘴唇来描述的一个动作,它可以带给自己或他人乐趣。因此,laugh 可以表示为下面的格框架(图 7.6)。

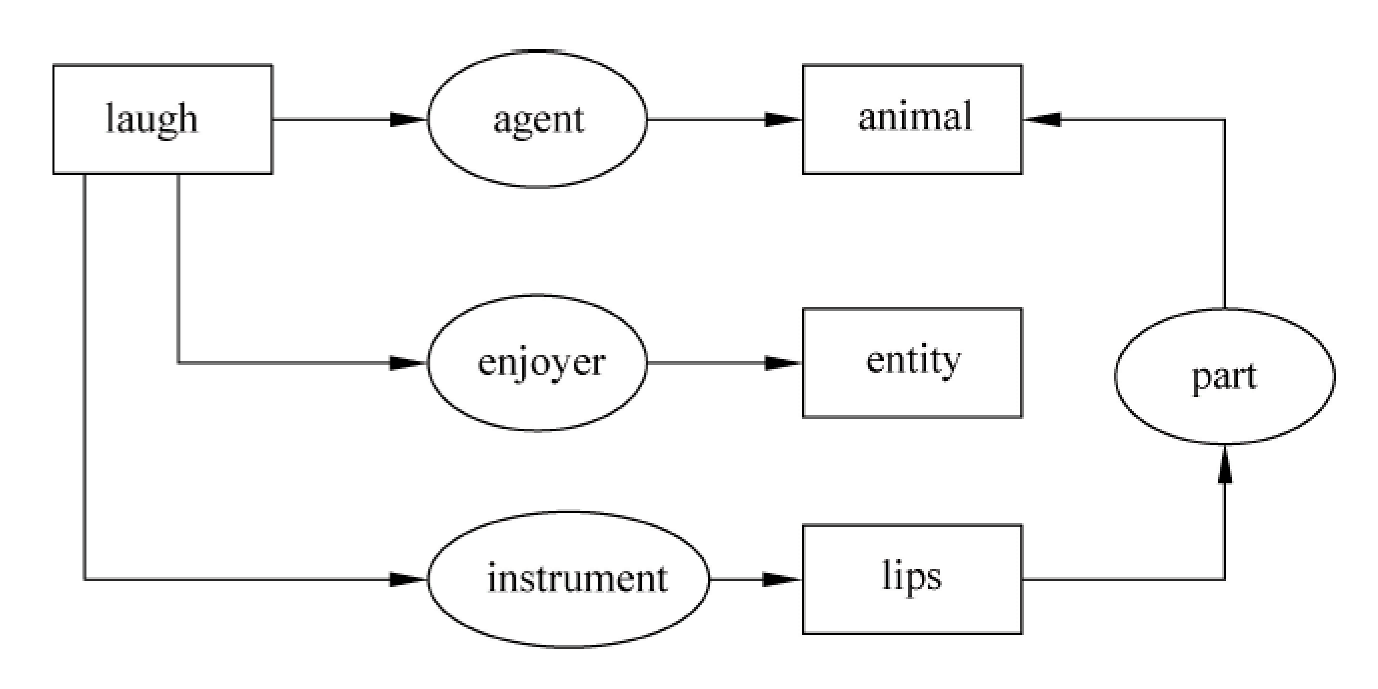



图 7.6　动词 laugh 的格框架

在图 7.6 中,矩形表示世界的描述,两个矩形之间的关系用椭圆表示。为了对 ATN 进行语义解释,需要指出如下几点。

(1) 当从 ATN 中的句子 S 开始分析时,需要确定名词短语和动词短语以得到名词和动

词的格框架表示。将名词和对应格框架中的主语(动作主体)关联在一起。

(2) 当处理名词短语时,需要确定名词;确定冠词的数特征(单数还是复数),并将动作的制造者和名词相关联。

(3) 当处理动词短语时,需要确定动词。如果动词是及物的,则找到其对应的名词短语,并说明它为动词的施加对象。

(4) 当处理动词时,检索它的格框架。

(5) 当处理名词时,检索它的格框架。

格文法是一种有效的语义分析方法,它有助于删除句法分析的歧义性,并且易于使用。格表示易于用语义网络表示法描述,从而多个句子的格表示相互关联形成大的语义网络,以便开发句子间的关系,理解多句构成的上下文,并用于回答问题。

## 7.6 概念依存理论

1972 年香克提出了概念依存理论[648],作为表示短语和句子的意思,为计算机提供常识知识以利于推理,从而达到对语言的自动理解。概念依存理论的基本原理如下。

(1) 对于任何两个意思相同的句子,不管语言,都该仅有一种概念依存意思的表示。

(2) 概念依存表示由非常少量的语义原构成,语义原包括原动作和原状态(与属性值有关)。

(3) 隐式句子中的任何信息必须形成表示那个句子意思的显式表示。

概念依存理论有如下三个层面。

(1) 概念依存层面→动作基元,包括:

① 物理世界的基本动作={抓 GRASP,移动 MOVE,传递 TRANS,去 GO,推 PROPEL,吸收 INGEST,撞击 HIT}。

② 精神世界的基本动作={心传 MTRANS,概念化 CONCEPTUALIZE,心建 MBUILD}。

③ 手段或工具的基本动作={闻 SMELL,看 LOOK-AT,听 LISTEN-TO,说 SPEAK}。

(2) 剧本→描写常见场景中的一些基本固定的成套动作(由动作基元构成)。

(3) 计划→其每一步由剧本构成。

下面,我们来介绍香克的概念依存关系,他把概念分为下列范畴。

(1) PP:一种概念名词,只用于物理对象,也叫图像生成者。例如人物、物体等都是 PP,还包括自然界的风雨雷电和思维着的人类大脑(把大脑看成一个产生式系统)。

(2) PA:物理对象的属性,它和它的值合在一起描述物理对象。

(3) ACT:一个物理对象对另一个物理对象施行的动作,也可能是一个物理对象自身的动作,包括物理动作和精神动作(如批评)。

(4) LOC:一个绝对位置(按“宇宙坐标”确定),或相对位置(相对于一个物理对象)。

(5) TIME:一个时间点或时间片,也分绝对或相对时间两种。

(6) AA:一个动作(ACT)的属性。

(7) VAL:各类属性的值。

香克采用下列方法形成新的概念体(Conceptualization)。

(1) 一个演员(能动的物理对象),加上一个动作(ACT)。

(2) 上述概念加上任选的下列修饰。

① 一个对象(若 ACT 为物理动作,则为一个物理对象,若 ACT 为精神动作,则为另一个概念)。

② 一个地点或一个接收者(如 ACT 发生在两个物理对象之间,表示有某个物理对象或概念体传到了另一个物理对象那里。如 ACT 发生在两个地点之间,表示对象的新地点)。

③ 一个手段(本身也是一个概念)。

(3) 一个对象加上此对象的某一属性的值。

(4) 概念和概念之间以某种方式组合起来,形成新的概念,例如,用因果关系组合起来。

本来,香克的目标是要把所有的概念都原子化,但事实上,他只做了对动作(ACT)的原子化。他将 ACT 分为如下 11 种。

(1) PROPEL:应用物理力量于一对象,包括推、拉、打、踢等。

(2) GRASP:一个演员抓起一个物理对象。

(3) MOVE:演员身体的一部分变换空间位置,如抬手、踢腿、站起、坐下等。

(4) PTRANS:物理对象变换位置,如走进、跑出、上楼、跳水等。

(5) ATRANS:抽象关系的改变,如传递(持有关系改变)、赠送(所有关系改变)、革命(统治关系改变)等。

(6) ATTEND:用某个感觉器官获取信息,如用目光搜索、竖起耳朵听等。

(7) INGEST:演员把某个东西吸入体内,如吃、喝、服药等。

(8) EXPEL:演员把某个东西送出体外,如呕吐、落泪、便溺、吐痰等。

(9) SPEAK:演员产生一种声音,包括唱歌、奏乐、号啕抽泣、尖叫等。

(10) MTRANS:信息的传递,例如交谈、讨论、打电话等。

(11) MBUILD:由旧信息形成新信息,如怒从心头起;恶向胆边生;眉头一皱,计上心来之类。

在定义这十一种原子动作时,香克有一个基本的思想,这些原子概念主要地不是用于表示动作本身,而是表示动作的结果,并且是本质的结果,因此也可以认为是这些概念的推理,例如:"X 通过 ATRANS 把 Y 从 W 处转到 Z 处"包含着如下推论。

(1) Y 原来在 W 处。

(2) Y 现在到了 Z 处(不再在 W 处)。

(3) 通过 ATRANS 实现了 X 的某种目的。

(4) 如果 Y 是一种好的东西,则意味着事情向有利于 Z,而不利于 W 的方向变化,否则相反。

(5) 如果 Y 是一种好的东西,则意味着 X 作此动作是为了 Z 的利益,否则相反。

一类重要的句子是因果链,香克和他们的同事制订了一些关于用于概念依存理论的一些规则。五种重要规则如下。

(1) 动作可以导致状态的改变。

(2) 状态可以启动动作。

(3) 状态可以消除动作。

(4) 状态(或者动作)可以启动精神事件。

(5) 精神事件可以是动作的原因。

这些是关于世界的知识的基本部分，概念依存包括每种（和组合）叫做因果连接的速记表示。在概念依存理论中，隐式句子中的任何信息必须形成表示那个句子意思的显式表示。例如句子 John eats the ice cream with a spoon（John 用匙吃冰淇淋）的概念依存表示为图 7.7。图中 D 和 I 矢量分别表示方向和使用说明依赖。注意，这个例子中，嘴是作为概念化部分进入图中，即使它没有出现在原来句子中。这是概念依存和句子语法分析产生的导出树之间的基本差别。

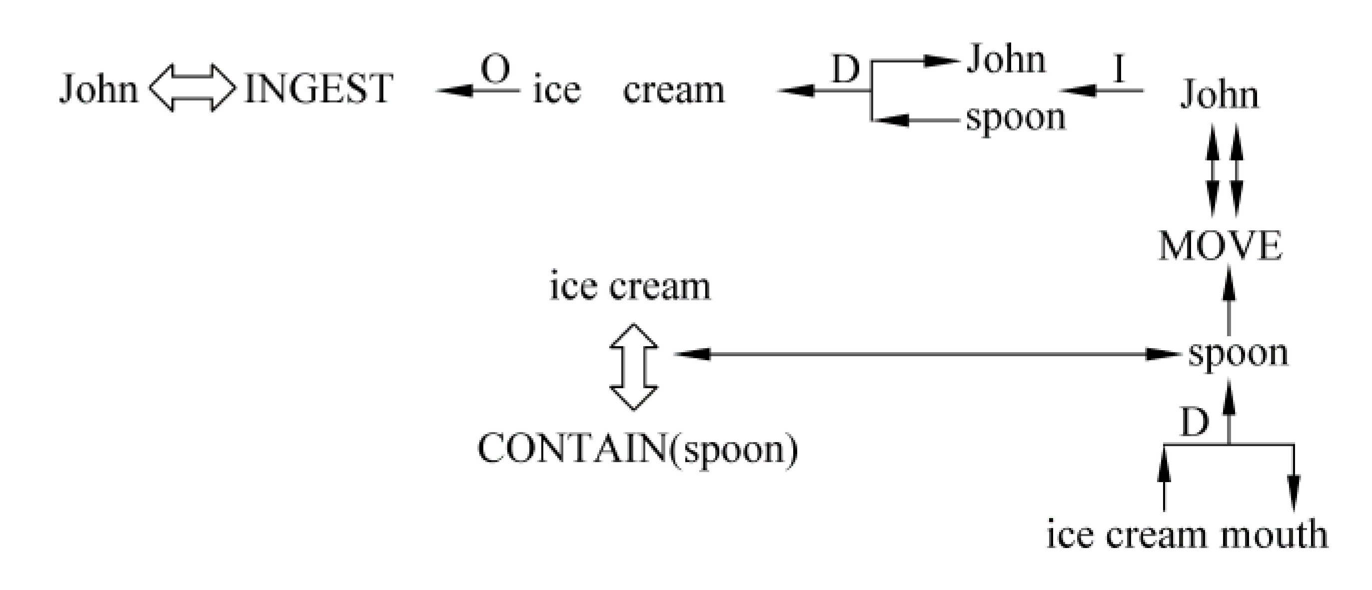



图 7.7 隐式信息的表示

# 7.7 语言理解

## 7.7.1 概述

语言理解，是指人们借助于听觉或视觉的语言材料，在头脑中构建意义的一种主动、积极的过程。它可以揭示语言材料所蕴含的意义。美国认知心理学家奥尔森（G M Olsen）提出语言理解的判别标准如下。

（1）能成功地回答语言材料中的有关问题，就是说，回答问题的能力是理解语言的一个标准。

（2）在给予大量材料之后，有做出摘要的能力。

（3）能够用自己的语言，即用不同的词语来复述这个材料。

（4）从一种语言转译到另一种语言。

自然语言信息处理研究在电子计算机问世之初就开始了，并于 20 世纪 50 年代初开展了机器翻译试验。20 世纪 70 年代，随着认知科学的兴盛，研究者相继提出了语义网络、概念依存理论、格框架等语义表示理论。这些语法和语义理论经过各自的发展，逐渐开始趋于相互结合。到 20 世纪 80 年代一批新的语法理论脱颖而出，具有代表性的有词汇功能语法（LFG）、功能合一语法（FUG）和广义短语结构语法（GPSG）等。这些基于规则的分析方法可以称之为自然语言处理中的“理性主义”。理性主义的基本出发点是追求完美，企图以思辨去百分之百地解决问题。美国著名的语言学家乔姆斯基在 20 世纪 60 年代提出标准理论，70 年代的扩展标准理论，80 年代的管辖与约束理论（Government and Binding Theory），以至 90 年代的最简方案（Minimalist Program），一直进行普遍语法（Universal Grammar）的研究[159,619]。理性主义所追求的目标是不断地抽象，在语言认知或者纯粹的语言学理论研究中，找到类似元素周期表的一种跨越不同语言的语法通则。现有的手段虽然基本上掌握

了单个句子的分析技术,但是还很难覆盖全面的语言现象,特别是对于整个段落或篇章的理解还很困难。

与"理性主义"相对的是"经验主义"的研究思路,主要是指针对大规模语料库的研究。语料库是大量文本的集合。计算机出现后,语料可以被方便地存储起来,利用计算机查找也很容易。随着电子出版物的出现,采集语料也不再困难。最早于 20 世纪 60 年代编制的 Brown 和 LOB 两个计算机语料库,分别具有 100 万词汇的规模。进入 20 世纪 90 年代可以轻易列举出的语料库有几十个之多,像 DCI、ECI、ICAME、BNC、LDC、CLR 等,其规模最高达到 $10^9$ 数量级。

对语料库的研究分成 3 个方面:工具软件的开发、语料库的标注、基于语料库的语言分析方法。采集到以后未经处理的生语料不能直接提供有关语言的各种知识,只有通过词法、句法、语义等多层次的加工才能使知识获取成为可能。加工的方式就是在语料中标注各种记号,标注的内容包括每个词的词性、语义项、短语结构、句型和句间关系等。随着标注程度的加深语料库逐渐熟化,成为一个分布的、统计意义上的知识源。利用这个知识源可以进行许多语言分析工作,如根据从已标注语料中总结出的频度规律可以给新文本逐词标注词性,划分句子成分等。

语料库提供的知识是用统计强度表示的,而不是确定性的,随着规模的扩大,旨在覆盖全面的语言现象。但是对于语言中基本的确定性的规则仍然用统计强度的大小去判断,这与人们的常识相违背。这种"经验主义"研究中的不足要靠"理性主义"的方法来弥补。两类方法的融合也正是当前自然语言处理发展的趋势。

自然语言理解系统的发展可以分为两个阶段:第一代系统建立在对词类和词序分析的基础之上,分析中经常使用统计方法;第二代系统则开始引进语义甚至语用和语境的因素,统计技术处于次要地位。第一代自然语言理解系统又可分为四种类型。

### 1. 特殊格式系统

早期的自然语言理解系统大多数是特殊格式系统,根据人机对话内容的特点,采用特殊的格式来进行人机对话。1963 年,林德赛(R Lindsay)在美国卡内基技术学院用 IPL-V 表处理语言设计了 SAD—SAM 系统,就采用了特殊格式来进行关于亲属关系方面的人机对话,系统内建立了一个关于亲属关系的数据库,可接收关于亲属关系方面的问题的英语句子提问,用英语作出回答。1968 年,波布洛(D Bobrow)在美国麻省理工学院设计了 STUDENT 系统,这个系统把高中代数应用题中的英语句子归纳为一些基本模式,由计算机来理解这些应用题中的英语句子,列出方程求解,并给出答案。

### 2. 以文本为基础的系统

某些研究者不满意在特殊格式系统中的种种格式限制,因为就一个专门领域来说,最方便的还是使用不受特殊格式结构限制的系统来进行人机对话,这就出现了以文本为基础的系统,1966 年西姆蒙(R F Simmons)、布尔格(J F Burger)和龙格(R E Long)设计的 PROTOSYNTHEX—Ⅰ系统,就是以文本信息的存储和检索方式工作的。

### 3. 有限逻辑系统

有限逻辑系统进一步改进了以文本为基础的系统。在这种系统中，自然语言的句子以某种更加形式化的记号来替代，这些记号自成一个有限逻辑系统，可以进行某些推理。1968年，拉菲尔（B Raphael）在美国麻省理工学院用 LISP 语言建立了 SIR 系统，针对英语提出了 24 个匹配模式，把输入的英语句子与这些模式相匹配，从而识别输入句子的结构，在从存储知识的数据库到回答问题的过程中，可以处理人们对话中常用的一些概念，如集合的包含关系、空间关系等等，并可进行简单逻辑推理，并能在对话中进行学习，记住已学过的知识，从事一些初步的智能活动。

### 4. 一般演绎系统

一般演绎系统使用某些标准数学符号（如谓词演算符号）来表达信息。逻辑学家们在定理证明工作上取得的全部成就，就可以用来作为建立有效的演绎系统的根据，从而能够把任何一个问题用定理证明的方式表达出来，并实际地演绎出所需要的信息，用自然语言作出回答。一般演绎系统可以表达那些在有限逻辑系统中不容易表达出来的复杂信息，从而进一步提高了自然语言理解系统的能力。1968—1969 年，格林和拉菲尔建立的的 QA2、QA3 系统，采用谓词演算的方式和格式化的数据来进行演绎推理，解答问题，并用英语作出回答，这是一般演绎系统的典型代表。

1970 年以来，出现了第二代自然语言理解系统，这些系统绝大多数是程序演绎系统，大量地进行语义、语境以至语用的分析。其中比较有名的系统是 LUNAR 系统、SHRDLU 系统、MARGIE 系统、SAM 系统、PAM 系统。

LUNAR 系统是伍兹（W Woods）于 1972 年设计的一个自然语言情报检索系统。这个系统采用形式提问语言来表示所提问的语义，从而对提问的句子作出语义解释，最后把形式提问语言执行于数据库，产生出对问题的回答。

SHRDLU 系统是维诺格拉德（T Winograd）于 1972 年在美国麻省理工学院建立了一个用自然语言指挥机器人动作的系统。该系统把句法分析、语义分析、逻辑推理结合起来，大大地增强了系统在语言分析方面的功能。该系统对话的对象是一个具有简单的“手”和“眼”的玩具机器人，它可以操作放在桌子上的具有不同颜色、尺寸和形状的玩具积木，如立方体、棱锥体、盒子等，机器人能够根据操作人员的命令把这些积木捡起来，移动它们去搭成新的积木结构，在人机对话过程中，操作人员能获得他发给机器人的各种视觉反馈，实时地观察机器人理解语言、执行命令的情况。在电视屏幕上还可以显示出这个机器人的模拟形象以及它同一个真正的活人在电传机上自由地用英语对话的生动情景。

MARGIE 系统是香克于 1975 年在美国斯坦福人工智能实验室研制出来的。该系统的目的在于提供一个自然语言理解的直观模型。系统首先把英语句子转换为概念依存表达式，然后根据系统中有关信息进行推理，从概念依存表达式中推演出大量的事实。由于人们在理解句子时，总要牵涉到比句子的外部表达多得多的内容，因此，该系统的推理有 16 种类型，如原因、效应、说明、功能等，最后，把推理的结果转换成英语输出。

PAM 系统是威林斯基（R Wilensky）于 1978 年在美国耶鲁大学建立的另一个理解故事的系统。PAM 系统也能解释故事情节，回答问题，进行推论，作出摘要。它除了“脚本”中

的事件序列之外,还提出了“规划”(Plan)作为理解故事的基础。所谓“规划”,就是故事中的人物为实现其目的所要采取的手段。如果要通过“规划”来理解故事,就要找出人物的目的以及为完成这个目的所采取的行动。系统中设有一个“规划库”(Plan Box),存储着有关各种目的以及各种手段的信息。这样,在理解故事时,只要求出故事中有关情节与规划库中存储的信息相重合的部分,就可以理解到这个故事的目的是什么。当在一个一个的故事情节与脚本匹配出现障碍时,由于“规划库”中可提供关于一般目的的信息,就不致造成故事理解的失败。例如,营救一个被暴徒抢走的人,在“营救”这个总目的项下列若干个子目的,包括到达暴徒的巢穴以及杀死暴徒的各种方法,就可以预期下一步的行为。同时能根据主题来推论目的。例如,输入故事:“约翰爱玛丽。玛丽被暴徒抢走了。”PAM 系统即可预期约翰要采取行动营救玛丽。故事中虽然没有这样的内容,但是,根据规划库中的“爱情主题”,可以推出“约翰要采取行动营救玛丽”的情节。

上述的系统都是书面的自然语言理解系统,输入输出都是用书面文字。口头的自然语言理解系统,还牵涉到语音识别、语音合成等复杂的技术,显然是更加困难的课题,口头自然语言理解系统的研究近年来也有进展。

## 7.7.2 基于规则的分析方法

从语言学和认知学的观念出发,建立一组语言学规则,使机器可以按照这组规则来正确理解它面对的自然语言。基于规则的方法是一种理论化的方法,在理想条件下,规则形成完备系统,能够覆盖所有语言现象,于是利用基于规则的方法就可以解释和理解一切语言问题。

自然语言理解系统都不同程度地涉及句法(Syntax)、语义学(Semantic)和语用学(Pragmatics)。句法是把词联结成短语、子句和句子的规则,句法分析是上述三个领域中迄今解决得最好的一个。大多数自然语言理解系统都包含一个句法分析程序、生成句法树(见图 7.8)之类的表示来反映输入语句的句法结构,以备进一步分析。图 7.8 给出了“事实证明张三是正确的”的句法树。

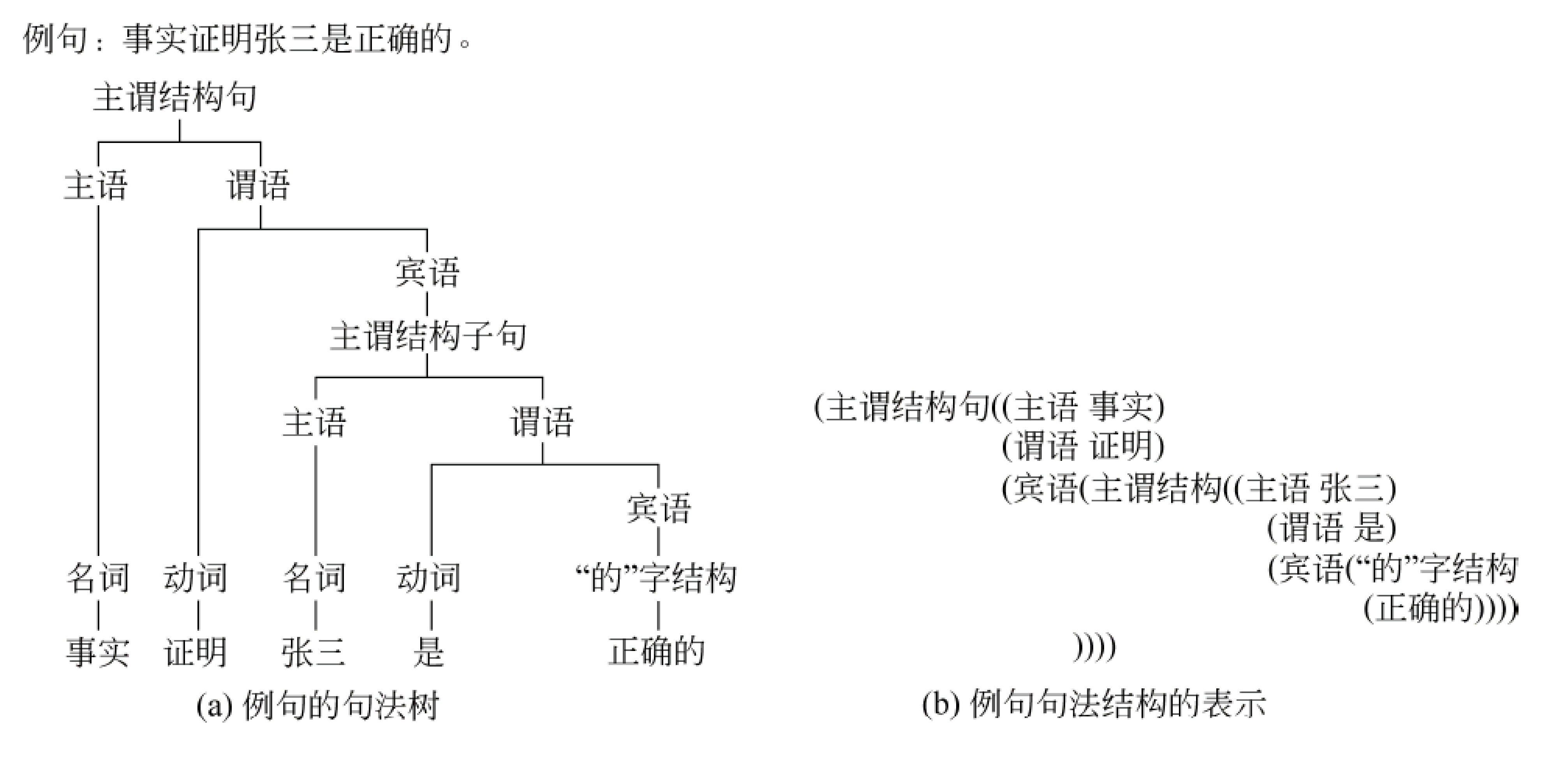



图 7.8 两种句法结构表示法

考虑到一些句法歧义的句子的存在，考虑到许多词在不同的语境(Context)中往往可以充当不同的词类，所以单纯依靠句法分析还往往不能获得正确的句法结构信息。因此有必要借助于某种形式的语义学分析。语义学考虑的是词义以及由词组成的短语、子句和语句所表达的概念。例如：

(1) 他在家。(“在”是动词)

(2) 他在家睡觉。(“在”是介词)

(3) 他在吃饭。(“在”是副词)

同一个“在”字，在不同的语境中可以分别充当不同的词类，而且含义也不同。这些例子可以说明即使在句法分析的过程中，为了尽快地获得正确的分析，往往需要某些语义信息，甚至外部世界知识的干预。在对句法与语义学分析问题上，目前大体上有以下两种不同的做法。

(1) 将句法分析与语义学分析分离的串行处理(如图 7.9(a)所示)。传统的语言学家主张把句法分析和语义分析完全分离开来。但许多著名的自然语言理解系统，如维诺格拉德的 SHRDLU 系统等，都允许在对输入语句进行句法分析的过程中调用语义学的解释因函数来辅助分析(如图 7.9(a)中的虚线所示)。尽管如此，它们都将产生某种形式的句法树来作为句法分析的结果。

(2) 句法与语义学的一体化处理方案，如图 7.8(b)所示。这是以耶鲁大学教授香克为代表的人工智能学派多年来竭力提倡的一种处理方案。这种方案的特点是取消了相对独立的句法分析模块，因而也不再生成反映输入语句句法结构的中间结果。他们的指导思想是尽可能早地在分析中综合引用包括词法句法、语义学、语境和世界知识在内的各种知识源。他们在分析中不是完全不要句法知识，只是不过分依赖于句法分析而已。不少心理学家也曾论证这种一体化的分析方案更接近人对语言的理解机制。这种方案的代表作是该学派的 ELI 和 CA 等英语分析系统。

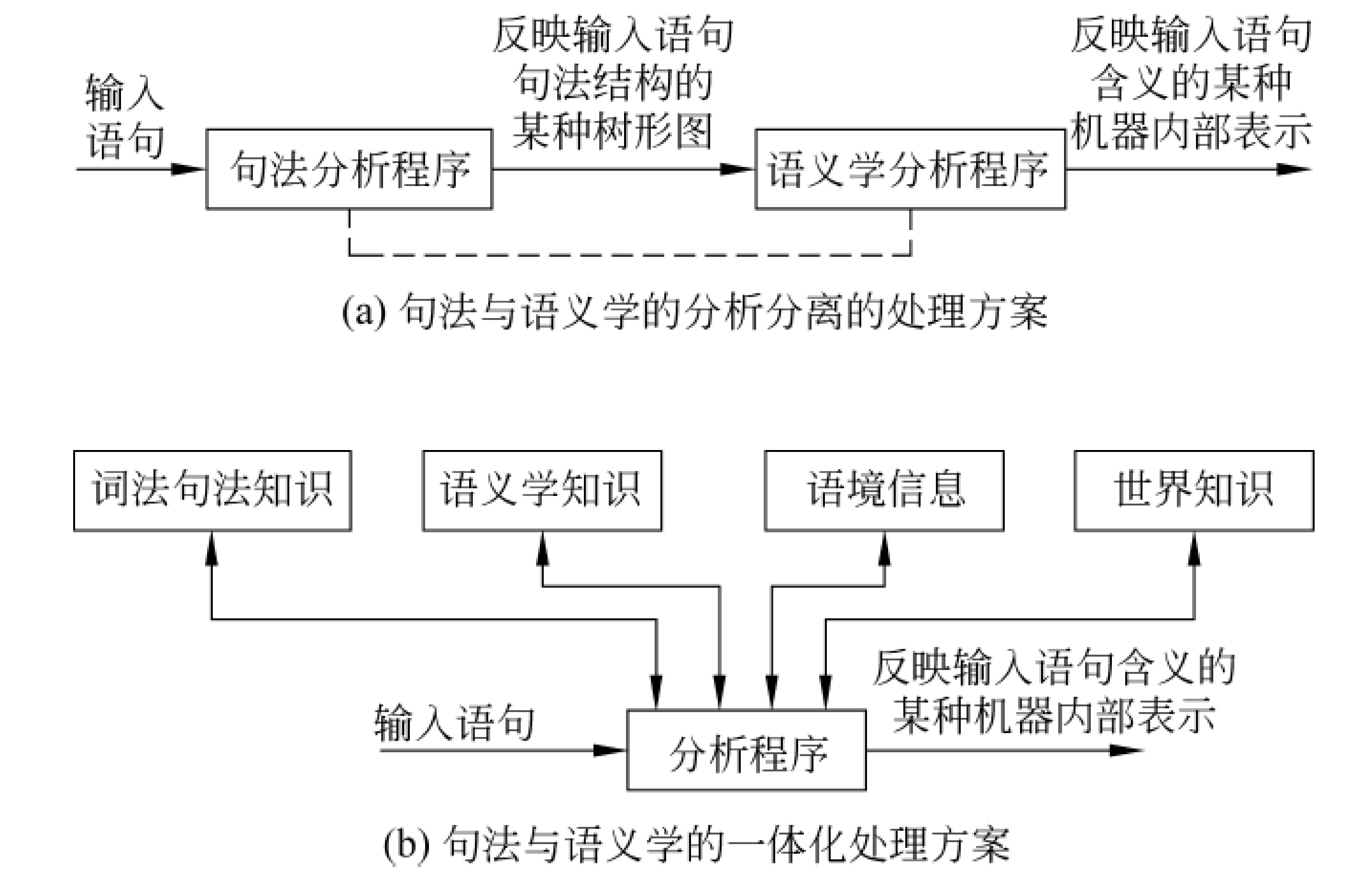



图 7.9　自然语言分析系统的两种方案

20 世纪 80 年代以来，国内外自然语言处理领域中对语义知识工程进行了研究。各语义知识库均以语义关系为重点描写内容；语义知识范畴具有明显的相对性特点；语义知识

主要是作为约束条件,在计算机对语言形式做各种变换操作时发挥作用;重视通过系统的语言形式变换手段来界定语义范畴,提取语义约束条件。由此得到的语义知识,能更好更直接地为自然语言处理服务。

把什么人写的句子,以及在什么地点(场合)、什么时间写的等等因素考虑进来,以便对句子作出更全面的解释,这就是语用学的任务之一。这种分析显然要求系统拥有更广泛的语境信息和世界知识。1972 年,维诺格拉德将语言学方法和推理方法结合,恰当处理了语法、语义和语用学的相互作用,在 PDP10 计算机上成功地开发了自然语言处理系统 SHRDLU。它是一个人类语言理解的一种比较有生命力的理论模型,引起了很多研究者的兴趣[813]。

这个系统包括:一个分析程序,一部英语的系统语法,一个语义分析程序,以及一个问题求解器。系统用 LISP 语言和 MICRO-PLANNER 语言写成,后者是一种基于 LISP 的程序语言。系统的设计建立在这样一种信念的基础上,即为了理解语言,程序必须以一种整体的观念来处理句法、语义和推理。计算机系统只有能够理解它所讨论的主题,才能合理地研究语言、系统,给出关于一个特殊领域的详尽模型。而且还有一个关于它自身智力的简单模型,例如它能回忆和讨论它的计划和行动。系统中知识是以过程的方式表示的,而不是以规则表格或模式来表示的。它通过对于句法、语义和推理的专门过程来体现,由于每份知识都可以是一个过程。它便能够直接调用系统中任何其他知识,因此 SHRDLU 系统有能力达到当时前所未有的性能水平。

钟义信提出一种全信息概念。自然语言所表达的信息都是(各种)认识主体所表述的信息,当然属于认识论层次信息的范畴。这里所说的认识主体,一般而言可以是人,可以是各种生物,也可以是人造的机器系统。不过,最有意义的认识主体是人自己。从认识论的观点看,由于正常的认识主体通常都具有观察力、理解力、目的性三个基本特性,因此,作为主体所感知或所表述的"事物运动状态及其变化方式",也必然包括如下三个方面。

(1) 事物运动状态及其变化方式的形式方面,称为事物的语法信息。

(2) 事物运动状态及其变化方式的含义方面,称为事物的语义信息。

(3) 事物运动状态及其变化方式对于认识主体的目的而言的效用方面,称为事物的语用信息。而语法信息、语义信息和语用信息组成一个有机整体,则称为全信息[982]。图 7.10 给出了全信息概念的形象化解释。

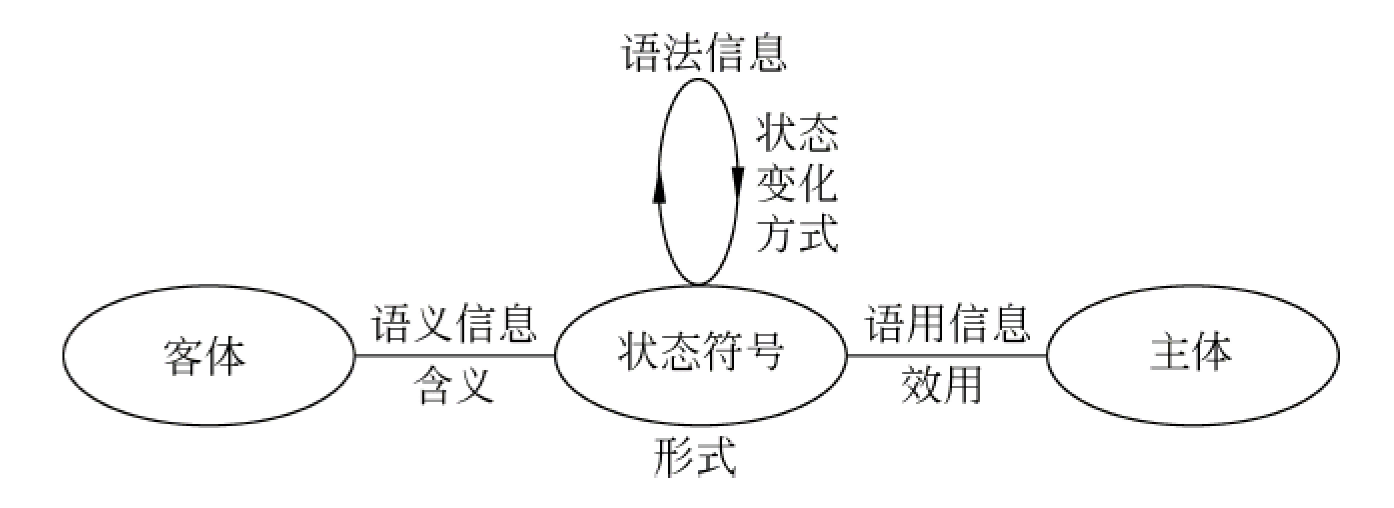



图 7.10　全信息概念图解

图 7.10 中,事物运动状态及状态变化方式的形式(图的中央部分)是事物的语法信息,它是事物运动状态及其变化方式的一种形式表现。一旦这种状态及其变化方式的形式与它相应的实际客体事物联系起来,它就会具有具体的实际含义(图的中央和左边的部分),这就

是语义信息，它不再是抽象的东西，而成为十分具体的东西；进一步，如果状态及其变化方式的形式以及它的含义一旦与特定的认识主体联系起来，它就会表现出对主体目的的效用(全图)，这就是语用信息。

由此可见，语法信息是一个抽象的信息层次；语义信息是语法信息与其相应客体互相关联的结果；语用信息则是语法信息、语义信息与认识主体相互关联的结果，因而是最具体的层次。语法信息和语义信息只与事物客体的情况有关，语用信息则还与主体的情形有关。可以看出，全信息概念是一个有机的体系。

### 7.7.3 基于语料的统计模型

语料库语言学研究自然语言机读文本的采集、存储、标注、检索、统计等。目的是通过对客观存在的大规模真实文本中的语言事实进行定量分析，支持语言学研究和鲁棒的自然语言处理系统的开发。应用领域包括语言文字的计量分析、语言知识获取、作品风格分析、词典编纂、全文检索系统、自然语言理解系统以及机器翻译系统等。

现代语料库语言学的起源可追溯到 20 世纪 50 年代美国布罗菲尔德(Leonard Bloomfield)后期的结构主义语言学时代，那时的语言学家在科学的实证主义和行为主义观点影响下，认为语料库是一个规模足够大的语言数据库。1990 年 8 月，在赫尔辛基召开的第 13 届国际计算机语言学大会上，大会组织者提出了处理大规模真实文本将是今后一个相当长的时期内的战略目标。为实现战略目标的转移，需要在理论、方法和工具等方面实行重大的改革。这种建立在大规模真实文本处理基础上的研究方法将自然语言处理的研究推向一个崭新的阶段。理解自然语言所需的各种知识恰恰蕴涵在大量的真实文本当中，通过对大量真实文本进行分析处理，可以从中获取理解自然语言所需的各种知识，建立相应知识库，从而实现以知识为基础的智能型自然语言理解系统。研究语言知识所用的真实文本称为语料，大量的真实文本即构成语料库。要想从语料库中获取理解语言所需的各种知识，就必须对语料库进行适当的处理与加工，使之由生语料变为有价值的熟语料。这样，就形成了一门新的学科语料库语言学(Corpus Linguistics)，可用于对自然语言理解进行研究。

如何建造语料库，并且语料库中都包括什么样的语义信息呢？这里以 WordNet 为例来说明。WordNet 是 1990 由 Princeton 大学的米勒(G A Miller)等人设计和构造的[501]。一部 WordNet 词典将近 95 600 个词形(51 500 单词和 44 100 搭配词)和 70 100 个词义，分为五类：名词、动词、形容词、副词和虚词，按语义而不是按词性来组织词汇信息。在 WordNet 词典中，名词有 57 000 个，含有 48 800 个同义词集，分成 25 类文件，平均深度 12 层。最高层为根概念，不含有固有名词。

知网(HowNet)是董振东研制的以汉语和英语的词语所代表的概念为描述对象，以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识知识库[859]。公布的中文信息结构库包含：

① 信息结构模式：271 个。

② 句法分布式：49 个。

③ 句法结构式：58 个。

④ 实例：11 000 词语。

⑤ 总字数：中文 60 000 字。

传统的词典通常是把各类不同的信息放入一个词汇单元中加以解释，包括拼音、读音、词形变化及派生词、词根、短语、时态变换的定义及说明、同义词、反义词、特殊用法注释，偶尔还有图示或插图，包含着相当可观的信息存储。但是，它还有一些不足，特别是用在自然语言理解时更显得不够。

例如，对于名词“树”，传统的词典一般解释为：一种大型的、木制的、多年生长的、具有明显树干的植物。基本上是上位词加上辨别特征。但是，这还不够，还缺少一些信息，例如：

(1) 它没有谈到树有根，有植物纤维壁组成的细胞，甚至也没有提及它们是生命的组织形式。但是在 WordNet 中，只要查一下它的上位词“植物”，就可以找到这些信息。

(2) 树的定义没有包括对等词的信息，不能推测其他种类的植物存在的可能性。

(3) 对于各种树都感兴趣的读者，除了查遍词典，没有别的办法。

(4) 每个人对树都有自己的认识，而词典的编撰者又没有将其写在树的定义中。如树包括树皮、树枝；树由种子生长而成，等等。

可以看出，普通词典中遗漏的信息中大部分是关于构造性的信息而不是事实性的信息。WordNet 是按一定结构组织起来的义类词典，主要特征表现如下。

(1) 整个名词组成一个继承关系。

WordNet 有着严格的层次关系，这样一个单词可以把它所有的前辈的一般性的上位词的信息都继承下来，可以提供全局性的语义关系，具有 IS-A 关系。

(2) 动词是一个语义网。

动词大概是最难以研究的词汇，在动词词典中，很少有真正的同义动词。表达动词的意义对任何词汇语言学来说都是困难的。WordNet 不做成分分析，而是进行关系分析。这一点是计算语言学界所热衷的课题，与以往的语义分析方法不同。这种关系讨论的是动词间的纵向关系，即词汇蕴涵关系。

WordNet 基于名词和动词以及其他词性的关系进行词类间的纵向分析，在国际计算语言学界有很大的影响。但是，它也有不足之处，如没有考虑横向关系。

从上面可以看出，传统的词典和语料库是不一样的。为了对自然语言理解进行研究，需要优先考虑的问题主要是大规模真实语料库的建设和大规模、信息丰富的机读词典的编制方法的研究。

大规模真实文本处理的数学方法主要是统计方法，大规模的经过不同深度加工的真实文本的语料库的建设是基于统计性质的基础，如果没有这样的语料库，统计方法只能是无源之水。从真实语料中获取自然语言的有关知识只能是一种理想。所以如何设计语料库，如何对生语料进行不同深度的加工以及加工语料的方法等，正是语料库语言学要深入进行研究的。语料库语言学研究的主要内容如下。

(1) 基本语料库的建设。

(2) 语料加工工具的研究，包括自动分词系统、词性标注系统、句法分析系统、义项标注系统和话语分析系统等。

(3) 通过语料加工建立起各种带有标注信息的“熟”语料库。

(4) 从语料库中获取语言知识的技术与方法。

为了使汉语语料库具有普遍性、实用性和时代性，作为共享的基础设施，提供自然语言处理的重要资源，应该建设由精加工语料库、基础语料库和网络语料库构成的多层次的汉语

语料库。那么建设语料库的研究重点将转向如何获取三个层次语料库的资源并有效的对它们进行利用。精加工语料库能够为各种语言研究提供好的、大量的语言处理规范和实例。基础语料库是一个覆盖面广、规模大的生语料库，通过它可以提供更翔实的语言分析数据。网络语料库是能够实现动态更新的语言资源，包含了很多新词语、新搭配和新用法，可以用于网络语言、新词语、流行语的跟踪研究，也可以用来观察语言的用法模式随时间的变化情况。可以用通过基于互联网的多层次汉语语料库克服传统语料库中数据稀疏和语料更新问题。在语料库规模上从底向上逐渐减少，但质量上(加工深度)逐渐提高。精加工语料库维持在 1000 万词次规模，而基础语料库在 1 亿词次以上比较合理，底层网络语料库是在线的开放的资源。

## 7.7.4　机器学习方法

机器学习是根据生理学、认知科学等对人类学习机理的了解，建立人类学习过程的计算模型或认知模型；发展各种学习理论和学习方法，研究通用的学习算法并进行理论上的分析；建立面向任务的具有特定应用的学习系统。这些研究目标相互影响相互促进。目前机器学习方法广泛用于语言信息处理中。

### 1. 文本分类

分类的目的是学会一个分类函数或分类模型(也常常称作分类器)，该模型能把数据库中的数据项映射到给定类别中的某一个。分类和回归都可用于预测。预测的目的是从利用历史数据纪录中自动推导出对给定数据的推广描述，从而能对未来数据进行预测。和回归方法不同的是，分类的输出是离散的类别值，而回归的输出则是连续数值。这里我们将不讨论回归方法。

要构造分类器，需要有一个训练样本数据集作为输入。训练集由一组数据库记录或元组构成，每个元组是一个由有关字段(又称属性或特征)值组成的特征向量，此外，训练样本还有一个类别标记。一个具体样本的形式可为：$(v_1, v_2, \cdots, v_n;\ c)$；其中 $v_i$ 表示字段值，$c$ 表示类别。

分类器的构造方法有统计方法、机器学习方法、神经网络方法等等。统计方法包括贝叶斯法和非参数法(近邻学习或基于案例的学习)，对应的知识表示则为判别函数和原型事例。机器学习方法包括决策树法和规则归纳法，前者对应的表示为决策树或判别树，后者则一般为产生式规则。神经网络方法主要是 BP 算法，它的模型表示是前向反馈神经网络模型(由代表神经元的节点和代表联接权值的边组成的一种体系结构)，BP 算法本质上是一种非线性判别函数。另外，最近又兴起了一种新的方法：粗糙集(Rough Set)，其知识表示是产生式规则。

### 2. 文本聚类

根据数据的不同特征，将其划分为不同的数据类。它的目的是使得属于同一类别的个体之间的距离尽可能的小，而不同类别上的个体间的距离尽可能的大。聚类方法包括统计方法、机器学习方法、神经网络方法和面向数据库的方法。

在统计方法中，聚类称聚类分析，它是多元数据分析的三大方法之一(其他两种是回归

分析和判别分析)。它主要研究基于几何距离的聚类,如欧式距离、明考斯基距离等。传统的统计聚类分析方法包括系统聚类法、分解法、加入法、动态聚类法、有序样品聚类、有重叠聚类和模糊聚类等。这种聚类方法是一种基于全局比较的聚类,它需要考察所有的个体才能决定类的划分;因此它要求所有的数据必须预先给定,而不能动态增加新的数据对象。聚类分析方法不具有线性的计算复杂度,难以适用于数据库非常大的情况。

在机器学习中聚类称作无监督或无教师归纳;因为和分类学习相比,分类学习的例子或数据对象有类别标记,而聚类的例子则没有标记,需要由聚类学习算法来自动确定。很多人工智能文献中,聚类也称概念聚类;因为这里的距离不再是统计方法中的几何距离,而是根据概念的描述来确定的。当聚类对象可以动态增加时,概念聚类则称是概念形成。

在神经网络中,有一类无监督学习方法:自组织神经网络方法;如 Kohonen 自组织特征映射网络、竞争学习网络等等。在数据挖掘领域里,见报道的神经网络聚类方法主要是自组织特征映射方法,IBM 在其发布的数据挖掘白皮书中就特别提到了使用此方法进行数据库聚类分割。

### 3. 基于案例的机器翻译

基于案例的机器翻译是日本学者长尾真于 20 世纪 90 年代初首先提出来的。该方法以基于案例推理(Case-based Reasoning,CBR)为理论基础。在 CBR 中,把当前所面临的问题或情况称为目标案例,而把记忆的问题或情况称为源案例。简单地讲,基于案例推理就是由目标案例的提示而获得记忆中的源案例,并由源案例来指导目标案例求解的一种策略。因此,基于实例的类比翻译其大致思路是:预先构造由双语对照的翻译单元对组成的语料库,然后翻译过程选择一个搜索和匹配算法,在语料库中寻找最优匹配单元对,最后根据例句的译文构造出当前所翻译单元的译文。

假设我们要翻译源语言文本 S,那么需要从事先已存好的双语语料库中找到与 S 相近的翻译实例 S′,再根据 S′的参考译文 T′来类比构造出 S 的译文 T。一般地,基于案例的机器翻译系统包括候选实例模式检索、语句相似度计算、双语词对齐和类比译文构造等几个步骤。如何根据源语言文本 S 找出其最相近的翻译实例 S′,是基于实例的翻译方法的关键问题。到目前为止,研究人员还没有找到一种简单通用的方法来计算句子之间的相似度。此外,评价句子相似度问题还需要许多人类工程学、语言心理学等知识来做保障。

基于案例的机器翻译方法几乎不需要对源语言进行分析和理解,只需要一个比较大的句对齐双语语料库,因此其知识获取相对比较容易,结合翻译记忆技术,系统能从零知识自举。如果语料库中有与被翻译句子相似的句子,那么基于案例的方法可以得到很好的译文,而且句子越相似,翻译效果越好,译文质量越高。基于案例的翻译方法,还有一个优点就是,实例模式的知识表示能够简洁方便地表示大量人类语言的歧义现象,而这种歧义现象是精确规则难以处理的。然而,基于案例的机器翻译方法其缺点也是显而易见的。当没有找到足够相似的句子时,翻译将宣布失败,这就要求语料库必须覆盖广泛的语言现象。

基于案例(记忆)的机器翻译软件塔多思(Trados)是由德国 Trados GmbH 公司开发的,2005 年 6 月被语言服务供应商 SDL 收购。2007 年版的塔多思由以下三个主要部分组成:Translator's Workbench,一个翻译记忆库创建和管理,可以在内置人工智能语言系统下根据模糊匹配原则自动翻译,可集成到 Microsoft Word 和 TagEditor 中;TagEditor,用

于编辑带标记的文本如 HTML 和 XML，可在 Microsoft 和 DTP 格式环境下自动过滤文本；MultiTerm 与 Translator's Workbench 配合使用的术语管理器。套件还有可替换的翻译环境，以及用于翻译图形接口的专用工具，翻译文档校对工具。塔多思的 2009 版还将塔多思和 SDLX 集成在一起。

卡耐基梅隆大学的 PanEBMT 系统，其语料库中包含了 280 多万条英法双语句对，尽管建立 PanEBMT 系统的研究人员同时还想了许多其他办法，但对于开放文本测试，PanEBMT 的翻译覆盖面只有 70%左右。因此，关于基于案例的机器翻译的研究，其主要的一个方面就是应着力于研究在规模相对小的案例模式库的条件下，如何提高翻译系统翻译的覆盖面，或者说在保持系统翻译效果的前提下，如何减小案例模式库的规模。为了达到这个目的，就需要从案例模式库中自动提取尽可能多的语言学知识，包括语法知识、词法知识和语义知识等，并研究其相应的知识表示等。

## 7.8 脑语言功能区

语言作为人脑的一种高级皮层功能备受关注。自 1861 年布洛卡发现 Broca 区后，神经语言学研究一直是脑科学研究中最热门的领域。一个多世纪以来，对语言的科学性研究已得出两条基本结论：一是脑的不同部位在语言中完成不同的功能；二是不同的脑区损伤产生不同的言语障碍。随着神经功能影像技术及电生理监测技术的进展，脑语言的功能区研究取得了较大的进展。

### 7.8.1 经典语言功能区

脑语言的功能区可分为运动性语言中枢和感觉性语言中枢，运动性语言中枢在额下回的后部（Brodmann 的 44、45 区，简写为 BA44、45），即 Broca 区。该区也称为前说话区，常常描述为额下回后 1/3。用于计划和执行说话，病变损伤该区会导致运动性失语，主要表现为口语表达障碍。辅助运动区 SMA 也称为上语言区，位于中央前回下肢运动区前方，后界为中央前沟，内侧界为扣带沟，外侧延伸至邻近的半球凸面，其前侧与外侧无明显界线。SMA 和初级运动区、运动前区扣带以及前额皮质背外侧、小脑、基底节、顶叶感觉联系区相互联系。这一复杂的解剖功能系统用于发动和控制运动功能和语言表达。进一步将 SMA 分为 SMA 前区和 SMA 固有区，分别参与复杂运动的准备和执行。

优势半球运动前区皮质 PMC 描述为初级运动皮质（Brodmann 4 区）、前方额叶无颗粒皮质区（Brodmann 6 区）。该区又分为两个亚区：腹侧 PMC（中央前回前部 Brodmann6 区的腹侧部分）和被侧 PMC（中央前沟前方的额上、中回后部 Brodmann6 区的被侧部）。研究发现腹侧 PMC 涉及发音，被侧 PMC 涉及命名。神经功能影像研究进一步支持优势半球 PMC 参与不同的语言成分，如阅读任务、复述单词及命名工具图片等。

在其下方额中回后部又有一书写中枢（BA8）。感觉性语言中枢可分为听觉性语言中枢和视觉性语言中枢，这两者之间无明确的界限，即 Wernicke 区。Wernicke 区也称为后说话区，一般指的是优势半球颞上回后部，但也有学者认为该区包括 Brodmann 41 和 42 区后方的颞上回、颞中回后部以及属于顶下小叶的缘上回和角回（BA22、39）。Wernicke 区与躯体

感觉(Brodmann5、7 区)、听(Brodman 41、42 区)和视(Brodmann18、19 区)区皮层有着密切的联系,用于分析和识别语言的感觉刺激。该区病变产生感觉性失语,表现为患者的语声调和语调均正常,与人交谈时不能理解别人说的话,答话语无伦次或答非所问,听者难于理解。

颞叶中部和内侧部是一复杂的多功能区,具有广泛的视觉和听觉功能。电刺激研究发现,左颞叶中部和内侧部在听觉语言中起重要作用。刺激该处能引起失语性异常,该处病变可引起与语言有关的轻微障碍,包括找词困难、命名缺陷等。

颞底语言区位于优势半球梭状回,距离颞极 3~7cm,是一个 Wernicke 区之外的独立区域。其下方的白质纤维束和 Wernicke 区下方的白质纤维束有直接联系。在电刺激研究中发现颞叶下部皮质的语言作用,主要是感觉性和表达性语言缺失。电刺激颞底语言区后 80%的患者出现命名和理解障碍。

随着研究的深入,相继发现了另外一些与语言有关的脑区。左侧颞叶下后部由于其来自大脑前和后动脉的双重供血,因此不易形成缺血损伤模型,被以往损伤灶模型研究所遗漏。后来发现这个区域与词汇的检索有关,被称为基底颞叶语言区[122]。

基底神经节具有语言的皮层下整合中枢的作用,它不仅调节运动、协调锥体系功能,同时支持条件反射、空间扣觉及注意转换等较简单的认知和记忆功能,且有证据表明,基底神经节可能参与和语言有关的启动效应、逻辑推理、语义处理、言语记忆及语法记忆等复杂的认知和记忆功能,有对语言过程进行加工、整理和协调的作用。其他研究还发现,除经典的语言功能区外,左侧顶上小叶、两侧梭状回、左侧枕下回、两侧枕中回、辅助运动区及额下回等都参与了语言的处理[392]。

从心理学的角度来看,语言需要的记忆方式主要有 3 种,即音韵、拼字和语义,即大脑中存在语言的音、形、义的加工。语言感觉传入可通过听、视和触觉(盲文),其传出途径可为发音、书写和绘图。采用不同的刺激方式可能会激活不同的功能区,如视、听和触觉功能区等;受试者的不同的反应方式又可激活一些脑区,如运动区、小脑等,这些区域的激活有时会干扰语言功能区的准确定位。目前,对语言的语义、音韵和拼字研究使得对脑的语言功能区又有了更精细的划分。

### 7.8.2 语义相关功能区

对词语的语义处理是人脑语言处理的一项基本活动,马默里(C J Mummery)等的研究发现,语义任务可激活广泛的区域,包括左侧颞中间后部和颞下部皮质前部[519]。

宾德尔(J R Binder)等进行声音传入刺激的研究[78]。刺激任务采用动物的名称,对照任务采用一般的声音刺激和休息状态,要求受试者用鼠标指示该动物是否本土生长和是否供人类使用。研究发现,对词汇语义的理解激活了许多的脑区,具有明显的左侧半球优势,包括两侧颞上回、左侧半球的大部分颞中回,且激活区向腹侧延伸到颞下回、梭状回和海马旁回、额上回和额下回大部、额中回的腹侧和背侧及扣带回的前部、角回以及胼胝体压部周围等区域。

对于额下回更深入的研究发现,音韵处理激活左侧额下回的背侧(BA44,45),而语义处理激活左侧额下回的脑室面(BA45,47)。

### 7.8.3 音韵相关功能区

对语言的处理中要把某一方面的影响，如语义或音位的因素完全排除是很困难的。行为学研究也表明，在对汉字和英语单词进行语义处理时不可避免地要进行音位处理。因此音位和语义的激活脑区有很多重叠。

海姆(S Heim)等采用事件相关方法的研究发现，Broca 区(BA44)的上部、左侧额叶(BA45，46)、颞叶(颞上回后部)在音韵任务时有显著的激活，而且这种激活同时表现在音韵的理解和产生两方面[314]。最近的研究发现，左侧颞上回和颞中回在识别语言和非语言的音韵中有明显的激活，并且以左侧为主。

进一步研究音韵结构中的2个基本成分音节和音素，发现它们分别激活左侧额中回和左侧额下回前部区域。

### 7.8.4 拼字相关功能区

布什(J R Booth)等运用视觉输入的词汇拼写任务和听觉输入的词汇音韵任务，观察正常人的语言功能区[92]。结果发现，梭状回(BA19，37)在处理文字的拼字形态时有明显激活；颞上回(BA22，42)则更多在处理文字的音位时激活。利用不同任务形式之间的交叉刺激，则发现缘上回和角回(BA40，39)负责拼字和音位之间的转换。

与英文或其他的西方字母文字相比，由于汉字的独特的方块空间结构，需要更多的关于字符拼写的处理。根据这样的理论，研究发现左侧颞中回(BA9)在汉字的识别上表现更多的激活，推断此脑区可能负责协调和整合汉字的文字形态。

### 7.8.5 双语者脑语言功能区

人类大脑内不同区域是否代表不同的语言？对这个问题的研究是语言功能成像的一个热点。许多学者把双语或多语个体作为受试者，观察其不同语言的激活表现。虽然得到的结果各有不同，但大多数学者认为，双语者的母语和第二语言有很多重叠的脑激活区域，包括 Wernicke 区和 Broca 区。例如，汉语和英语双语者对于视觉刺激在左额和左颞可见高度重叠的激活脑区；西班牙语和(西班牙)加泰罗尼亚语双语者在听故事时，在左颞叶和海马区可见重叠。西班牙和英语双语者在造词时也发现左额叶、颞叶和顶叶广泛的重叠。

除母语和第二语言区有很多重叠的脑激活区城外，经常发现第二语言区比母语激活的范围广，而且激活的强度也大。这种现象经常发生在不太流利的双语者，在流利的双语者则无这种表现，与第二语言学习开始的时间无关，而与第二语言使用的频率有关。

# >>第 8 章
# 学习

人类通过学习来提高和改进自己的能力。学习的基本机制是设法把成功的表现行为转移到另一种类似的新情况中去。人的认识能力和智慧才能就是在毕生的学习中逐步形成、发展和完善。任何具有智能的系统必须具备学习的能力。学习能力是学习的方法与技巧，是人类智能的根本特征。

## 8.1 概述

1983 年西蒙对学习下了一个比较好的定义："系统为了适应环境而产生的某种长远变化，这种变化使得系统在今后能够更有成效地完成同一或同类的工作。"学习是一个系统中所发生的变化，它可以是系统作业的长久性的改进，又可以是有机体在行为上的持久性的变化。在一个复杂的系统中，由学习引起的变化是多方面的，也就是说，在同一个系统中可能包含着不同形式的学习过程，它的不同部分会有不同的改进。人在学习中获得新的产生式，建立新的行为。

学习的原理是学习者必须知道最后的结果，即其行为是否能得到改善。最好他还能得到关于他的行为中哪些部分是满意的，哪些部分是不满意的信息。对于学习结果的肯定的知识本身就是一种报酬或鼓励，它能产生或加强学习动机。关于学习结果的信息和动机的共同作用在心理学中叫作强化，其关系如下：

强化 ＝ 结果的知识 ＋ 报酬
（信息）　（动机）

强化不一定是外在的，它也可以是内部的。强化可以是积极的，也可以是消极的。学习时必须有一个积极的学习动机。强化能给学习动机以支持。老师在教育中要注意学习材料的选择，以吸引学生的注意，激励他们的学习。学习材料太简单，学生的精力不容易集中，容易产生厌烦情绪；学习材料太复杂，学生不容易理解，也会产生疲劳。可见，在学习中影响学习动机的因素是多方面的，其中包括学习材料的性质和构成等。

作者提出了一种学习系统模型（见图 8.1）。椭圆形表示信息单元，而长方形表示处理单元。箭头表示学习系统中数据流的方向。

影响学习系统最重要的因素是提供系统信息的环境，特别是这种信息的水平和质量。环境对学习单元提供信息。学习单元利用这些信息改善知识库。执行单元利用知识库执行它的任务。最后，执行任务时所获得的信息可以反馈给学习单元。若是人的学习，则通过内省学习机产生学习的效用信息，反馈给学习单元。

一百多年来，心理学家在探讨学习理论的过程中，由于各自的哲学基础、理论背景、研究

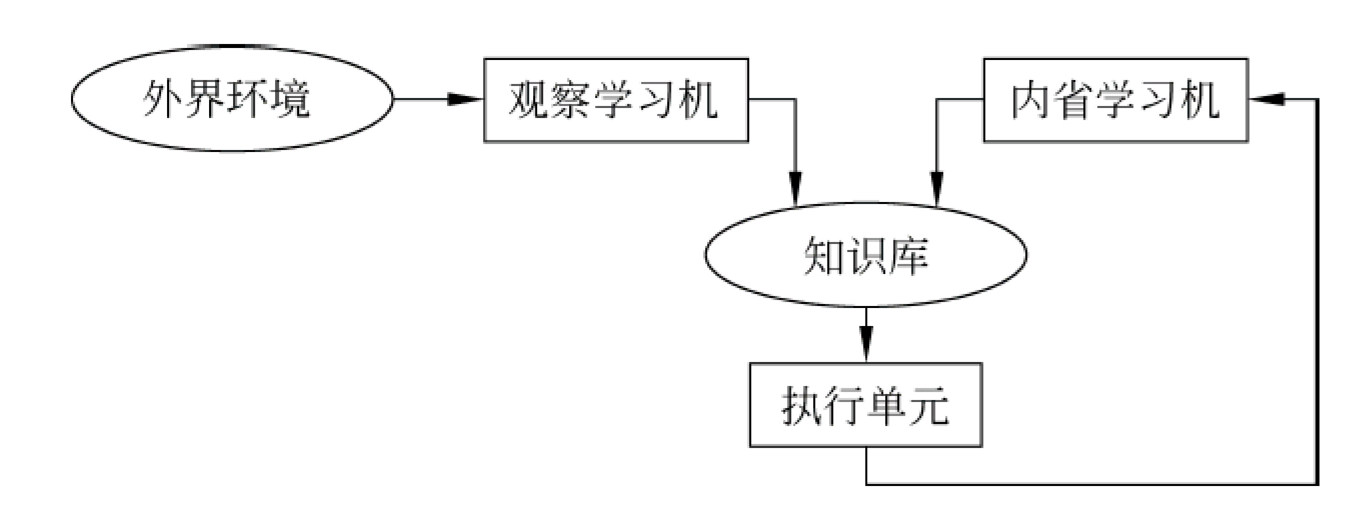



图 8.1　学习系统模型

手段的不同，自然形成了各种不同的理论观点，并形成了各种不同的理论派别，主要包括行为学派、认知学派和人本主义学派。

## 8.2　行为学习理论

有些心理学家用刺激与反应的关系，把学习解释为习惯的形成，认为通过练习使某一刺激与个体的某种反应建立一种前所未有的关系，此种刺激反应间联结的过程，就是学习。因此，此种理论被称为刺激反应论，或称为行为学派。行为学习理论强调可观察的行为，认为行为的多次愉快或痛苦的后果改变了个体的行为。巴甫洛夫经典条件反射学说、华生(J B Watson)的行为主义观点、桑代克(Thorndike)的联结主义、斯金纳(B F Skinner)的操作条件反射学说以及班图拉(A Bandura)的社会学习理论可作为行为派的代表学说。

另外有些心理学家不同意学习即习惯形成的看法，他们特别强调理解在学习过程中的作用。他们认为，学习是个体在其环境中对事物间关系认知的过程。因此，这种理论被称为认知论。

### 8.2.1　条件反射学习理论

俄国生理学家巴甫洛夫是经典条件反射学说的创立者[549]。巴甫洛夫在研究狗的消化生理现象时，把食物呈现在狗面前，并测量其唾液分泌。通常狗吃食物时才会分泌唾液。然而，巴甫洛夫偶然发现狗尚未吃到食物，只是听到送食物的饲养员的脚步声，便开始分泌唾液。巴甫洛夫没有放过这一现象。他开始做一个实验。先给狗听一个铃声，狗没有反应，然而在给狗铃声之后紧接着呈现食物，并经反复多次结合后，单独听铃声而没有食物，狗也“学会”了分泌唾液。铃声与无条件刺激(食物)的多次结合从一个中性刺激变成了一个条件性刺激，引起了分泌唾液的条件性反应，巴甫洛夫将这一现象称作条件反射，即经典条件反射。巴甫洛夫认为条件反射的生理机制是暂时神经联系的形成，并认为学习就是暂时神经联系的形成。

巴甫洛夫的经典条件反射学说的影响是巨大的。在俄国，以巴甫洛夫的经典条件反射学说为基础的理论在心理学界相当长的时间内曾占统治地位。在美国，行为派的心理学家华生、斯金纳等均接受到巴甫洛夫的条件反射学说的影响。

### 8.2.2　行为主义的学习理论

行为主义由美国心理学家华生于 1913 年创立[800]。该理论的特征有以下四点。

(1) 强调心理学是一门科学，因此在方法上重视实验、观察，在研究题材上只重视可观

察记录的外显行为。

(2) 解释构成行为的基础是个体表现于外的反应,而反应的形成与改变是由制约作用的历程。

(3) 重视环境对个体行为的影响,不承认个体自由意志的重要性,故而被认为是决定论。

(4) 在教育上主张奖励与惩罚兼施,不重视内发性的动机,强调外在控制的训练价值。

行为学派盛行在美国,影响扩及全世界,20 世纪 20 年代至 50 年代,期间四十多年,心理学界几乎全为行为主义的天下。行为主义也称为行为心理学。行为主义演变到后来,因对行为解释的观点不同,又有激进行为主义(Radical Behaviorism)与新行为主义(Neo-behav-iorism)之分。

华生是第一位将巴甫洛夫的经典条件反射学说作为学习的理论基础的美国心理学家。他主张一切行为都以经典条件反射学说为基础。他认为学习就是以一种刺激代替另一种刺激建立条件反射的过程,除了出生时具有的集中条件反射(如打喷嚏、膝跳反射)外,人类所有的行为都是通过条件反射建立新的刺激、反应联结(即 S-R 联结)而形成的。

华生运用条件反射的原理,做了一个婴儿恐惧形成的实验以证明他的观点。实验的对象是原来对兔子无任何恐惧的婴儿。在实验中,当兔子在婴儿面前出现时,同时发出一种可怕的声音,经多次重复后,婴儿见到兔子就会感到害怕,甚至会泛化到对任何有毛的东西感到恐惧。

### 8.2.3 联结学习理论

自 19 世纪末至 20 世纪初将近 50 年的时间内,桑代克的学习理论,在美国心理学界居于领导地位。桑代克是动物心理学研究的先驱,他从 1896 年开始,在哈佛大学用小鸡、猫、狗、鱼等动物做实验研究的对象,系统地研究动物的学习行为,从而提出了学习心理学中最早也最为完整的学习理论。通过科学的实验方法,他发现在学习环境中,个体的学习经由一种"尝试与错误偶然成功"的方式。在这种方式下,个体经过对刺激的多次反应,使两者间建立一种联结或结合。桑代克认为学习的实质在于形成情境与反应之间的联结,因此,这种学习理论被称为联结论。

情境(以 $S$ 代表)有时也叫刺激,包括外界情境和思想、情感等大脑内部情境。反应(以 $R$ 代表)包括"肌肉与腺体的活动"和"观念、意志、情感或态度"等内部反应。所谓联结,就是结合、关系、倾向,指的是某种情境只能唤起某种反应,而不能唤起其他反应的倾向。用"→"作为引起或导致的符号。联结的公式为:$S \rightarrow R$。

情境与反应之间是因果关系。它们之间是直接的联系,不需要任何中介。桑代克认为联结即本来(本能)的结合,是先天决定的原本趋向。他把联结的观点搬运到人类的学习上,认为人类所有的思想、行为和活动,都能分解为基本的单位刺激和反应的联结。人与动物学习的区别在于:"动物的学习过程全属盲目","无须观念为媒介",而人的学习是以观念为媒介,是有意识的。但二者的本质区别仅在于简单与复杂、联结数量的多少,动物学习的规律依然适合于人类的学习。

刺激与反应间的联结受以下三个原则的支配。

(1) 练习的多寡。

(2) 个体自身的准备状态。

(3) 反应后的效果。

这三个原则,就是桑代克的著名的学习三定律——练习律、准备律、效果律。练习律是指个体对某一刺激反应时,练习的次数愈多,则刺激与反应间的联结愈强。准备律即动机原则。动机是指引起个体活动,维持该种活动,并导致该种活动朝向某一目标进行的一种内在过程。效果律是联结论的核心,其主要内容在于强调刺激反应间联结的强弱要靠反应后的效果来决定。若反应后使个体获得满足的效果,则刺激反应间的联接加强;反之,若得到的是烦恼的效果,则刺激反应间的联结便减弱。

1930 年以后,桑代克对练习律和效果律做了修改,认为练习并不能无条件地增强情境与反应之间的联结的力量,而只有伴随着满意感时,练习才有作用。对于效果律,他认为烦恼对联结并无直接的削弱,只承认满意能增强联结。桑代克认为联结是通过尝试与错误的过程而建立的。学习的进程是一种渐进的、盲目的、尝试与错误的过程。在此过程中,随着错误反应的逐渐减少和正确反应的逐渐增加,而最终在刺激与反应之间形成牢固的联结。桑代克用不同的动物进行实验,结果相当一致,由此,他认为联结的形成是遵循着一定规律的。桑代克还提出学习的副律,包括:①选择反应律;②多重反应律;③定势律;④类化反应律;⑤联想转移律。

桑代克的学习理论是教育心理学史上第一个较为完整的学习理论。他运用实验而不是思辩的方法研究学习是一大进步。他的学习理论引起了有关学习理论的学术论争,推动了学习理论的发展。联结说的提出也有利于确立学习理论体系中的核心地位,相应地,也有利于教育心理学学科体系的建立,推动了教育心理学的发展。

联结说以本能作为学习的基础,以情境与反应的联结公式作为解释学习的最高原则,是遗传决定论和本能主义;它抹杀了人的学习的社会性,尤其是取消了人的学习的意识性和能动性,未能揭示人的学习的实质以及人的学习与动物学习的本质区别,是机械主义的。试误说以尝试和错误概括所有的学习过程,忽视了认知、观念、理解在学习过程中的作用,不符合学习的实际。但试误说直至今日仍被看成学习的一种形式,特别是在运动技能的学习和社会行为的学习中起着重要作用。桑代克提出的学习规律有些简单,不能完善到说明学习的根本规律,不过也有部分的真理性,即使现在来看,其中的一些规律对学习活动仍具有指导意义。

### 8.2.4　操作学习理论

操作学习理论是美国新行为主义心理学家斯金纳在《语言行为》一书中提出的言语学习理论。这一理论以对动物进行的操作性条件反射实验为基础,认为儿童获得言语主要靠后天学习,也与学习其他行为一样,是通过操作性条件反射来实现的。

1938 年斯金纳在特制的实验箱内研究了白鼠的学习。箱内装有一个杠杆,杠杆与传递事物的机械装置相连,只要杠杆一被压动,一颗食便滚进食盘。白鼠被放进箱内,自由活动,当它踏上杠杆时,有食丸放出,于是吃到食物。它再按压杠杆,食丸又滚出,反复几次,白鼠就学会了按压杠杆来取得食物的条件反射。斯金纳将这种条件反射叫做操作性条件反射。此外,斯金纳还做了鸽子啄圆窗反应的实验。

斯金纳认为条件反射有两种,即巴甫洛夫的经典性条件反射和操作性条件反射。巴甫洛夫的经典条件反射是应答性(或刺激性)条件反射过程,是先由已知刺激物引起的反应,是

强化物和刺激物相结合的过程,强化是为了加强刺激物的。斯金纳的操作性条件反射是反应型条件反射的过程,没有已知的刺激,是由有机体本身自发出现的反应,是强化物和反应相结合的过程,强化是为了增强反应的。

斯金纳认为一切行为都是由反射构成的,反射有两种,行为也必然有两种,即应答性行为和操作性行为。因此,学习也分为两种,即反射学习和操作学习。斯金纳更重视操作学习,他认为操作行为更能代表人在现实中的学习情况,认为人的学习几乎都是操作学习。因此,行为科学最有效的研究途径是研究操作行为的形成及其规律。

斯金纳认为强化是操作性行为形成的重要手段。强化在斯金纳的学习理论中占有极其重要的地位,是他学习理论的基石和核心,有人称他的学习理论为强化理论或强化说。操作学习的基本规律是：如果一个操作发生后,接着呈现一个强化刺激,则这个操作的强度(反应发生的概率)就增加。认为学习和行为的变化是强化的结果,控制强化就能控制行为。强化是塑造行为和保持行为强度的关键。塑造行为的过程就是学习过程。教育就是塑造行为。只要安排好强化程序,就可以随意地塑造人和动物的行为。

1953年秋天,斯金纳旁听了自己孩子所在小学四年级的课程。这就是他日后设计程序学习的开始。课堂教学是由一名教师讲解,许多儿童在这一位教师指导下进行学习。儿童存在着很大的个体差异,但教学却只能按照班里的平均水平进行。总之,班级教学很难按照每个儿童的个体差异进行教学。因此,他设计了适应个体差异教学的"教学机器"。

1954年,斯金纳在《学习科学与教学的艺术》(*The Science of Learning and the Art of Teaching*)一文中[726],根据他的强化理论,对传统教学进行了批评,指出：

(1) 传统教学在控制学生行为的手段上是消极的,多为负强化(如发脾气、惩罚、训斥等)。

(2) 行为和强化之间的时间间隔太长。

(3) 缺乏连续的强化程序。

(4) 强化太少。传统教学的最主要缺点就是强化太少。一个教师要对一个班几十名学生提供足够数量的强化机会是做不到的。

由此,斯金纳强力主张改变传统的班级教学,实行程序教学和机器教学。根据操作性条件反射原理把学习的内容编制成"程序"安装在机器上,学生通过机器上的程序显示进行学习。后来还发展了不用教学机器,只使用程序教材的程序学习。

程序学习的过程是将要学习的大问题分解成若干小问题,按一定顺序呈现给学生,要求学生一一回答,然后学生可得到反馈信息。问题相当于条件反射形成过程中的"刺激",学生的回答相当于"反应",反馈信息相当于"强化"。程序学习的关键是编制出好的程序。为此,斯金纳提出了编制程序的五条基本原理(原则)。

(1) 小步子原则：把学习的整体内容分解成由许多片段知识所构成的教材,把这些片段知识按难度逐渐增加排成序列,使学生循序渐进地学习。

(2) 积极反应原则：要使学生对所学内容作出积极的反应,否认"虽然没有表现出反应,但是的确明白"的观点。

(3) 及时强化(反馈)原则：对学生的反应要及时强化,使其获得反馈信息。

(4) 自定步调原则：学生根据自己的学习情况,自己确定学习的进度。

(5) 低的错误率：使学生尽可能每次都作出正确的反应,使错误率降到最低限度。斯金纳认为程序教学有如下优点：循序渐进；学习速度与学习能力一致；及时纠正学生的错

误,加速学习;利于提高学生学习的积极性;培养学生的自学能力和习惯。

斯金纳认为程序教学有如下优点:循序渐进;学习速度与学习能力一致;及时纠正学生的错误,加速学习;利于提高学生学习的积极性;培养学生的自学能力和习惯。程序学习并非尽善尽美。由于它主要是以掌握知识为目标的个体化学习方式,因此,人们对它的非议主要有三个方面:使学生学习比较刻板的知识;缺少班集体中的人际交往,不利于儿童社会化;忽视了教师的作用。

### 8.2.5　相近学习理论

格思里(E R Guthrie)承认刺激反应的联结是学习的必要条件,但却否定了增加在学习中的重要性。他认为学习的主要条件是刺激与反应之间时间上的相近,因此被称为相近论。格思里认为:“凡曾引起过某种反应的刺激再出现时,将会再引起同样的反应”。这一简单的原则,即代表他理论的全部。其中未谈到动机与增强,也未谈到重复练习的问题,认为构成学者,仅需刺激与反应的联结。

这种简单的原则,怎样用来解释复杂的学习现象呢?他认为任何一个刺激与反应之间,皆包含了很多小刺激和小反应。我们所看到的刺激反应,绝非单纯的一个刺激对一个反应,而是一连串刺激反应的两端。在一种技能学习中,所包含的动作越多,需要的练习也越多,如此才能把很多分立的刺激反应联结在一起。例如,投篮球是一组刺激反应的过程,其中包括球与手接触的部位、球传来时的力量、方向、球出手时的角度、离球篮的远近,有无敌方盯随等,构成一串极复杂的过程。

格思里从数学、哲学转向心理学,主要是由于受他的老师、哲学家辛格(E A Singer)的影响。在辛格看来,哲学上的许多问题都可以还原成行为的问题,并在行为的水平上加以分析。换言之,哲学问题可以通过实验的方法得到更好的解决。格思里早期的理论观点与华生和巴甫洛夫的学说很相似,但经过多年发展之后,形成了他自己独特的性质。格思里的主要代表著有:《学习心理学》(*The Psychology of Learning*)和《人类冲突心理学》(*The Psychology of Human Conflict*)。

到了 1935 年,格思里《学习心理学》一书出版了。在这本书里,他提出了有自己特色的学习理论。尽管他以后一直在不断地修正和完善其理论,甚至在他去世那年发表的《相近的联想》(*Association by Contiguity*)一文中还在修正他以往的观点,但他的理论框架和基本观点在 1935 年就已大致形成。下面就其主要论点做些介绍和分析。

格思里认为:一部好的学习心理学必须提出学习的规律,而且这种规律必须是容易理解和可以证实的。事实上,格思里一生中只提出了唯一的一条学习律,并用它来解释一切有关学习的问题。这就是刺激的某一组合,如果曾伴同过某一动作,那么,当这种刺激组合再次出现时,这一动作往往也会随之发生。直到 1959 年,格思里仍然在重述这一学习律,可见他对其效度坚信不移。

这条学习律不包括其他学习理论家所注重的强化或动机等变量,只强调学习即刺激与反应之间的联结。在格思里看来,当有机体对某种刺激作出某种反应时,便已充分达到了它们之间联结的力量。当这种刺激再次出现时,就会再次作出这种反应。因此,学习是一次性完成的。

这条学习律看上去有悖常理,且不容易理解。实际上,这是由于格思里对“刺激”与“反

应”这两个术语做了与众不同的解释而引起的。为了便于理解他的学习理论,这里有必要做些解释。

首先是对刺激的解释。人们通常把整个刺激情境看作是一种刺激。而格思里认为,在现实中,刺激情境是由时时刻刻在发生变化的大量刺激要素构成的。事实上,有机体在某一时刻只是对刺激情境中的某些刺激要素作出反应。所以,格思里的“刺激”概念,等同于环境中影响感觉接受器的那些特征。

其次,格思里对作为动作(Movement)的反应与作为行动(Act)的反应作了区分。前者是指个别的肌肉动作;后者是指由一系列肌肉动作构成的动作复合体。在他看来,行动是无法如实逼真地描述的。行动发生在整体反应之中,而整体反应是无限复杂的,同时又是与其他反应融合在一起的。因此,整体反应既没有名称,又无法描述。例如,投篮球不是一个单一的动作,而是由一连串肌肉动作组成的行动。格思里关心的是动作(即投篮时各种肌肉的动作)而不关心行动或行动的结果(即不关心整个投篮的过程和球是否投中)。格思里认为,正是这些动作与各组刺激要素形成的联结,才是学习的真正内涵。换言之,学习即是在刺激要素与肌肉动作之间形成的联结。

由此,格思里进一步认为,只要有机体对一组刺激要素作出一个动作,便在它们之间形成了联结,学习也就完成了。学习是一种全或无(All-or-None)的联结。如果有机体接下来做出另一个动作,那就说明有机体形成了一个新的联结,或者说发生了新的学习。这样,前一次的联结也就凋谢了。那么,有机体以后遇到类似刺激时,会作出哪一种反应呢?根据邻近的原理,有机体在某种刺激条件下最后作出的反应,是它以后遇到这种刺激时会作出什么反应的最佳预测者。由于格思里把他的学习律建立在刺激-反应联结的邻近上,也就是说,他认为在学习中起最主要作用的机制是暂时联系(Temporal Relatedness),即刺激与反应之间的紧密联系,因此被称为“邻近学习理论(Contiguity Theory of Learning)”。

动作-产生的刺激对于坚持行为主义立场的人来说具有重要意义,这在解释当有机体在不那么复杂的环境里从事各种复杂行为时表现得尤为明显。例如,一个人坐在书房里写作,外部情境并没有经历多大变化,但他仍然能一直坐在那里从事一系列复杂的写作活动。在这种情况下,光凭外部刺激显然难以对此做出解释,这就需要有引发行为的其他来源。格思里找到了。他认为,如果每一种反应都可以产生引发下一个反应的刺激,那么,在没有明显的外部刺激变化的情况下也可以维持行为了。这样,“这种内隐的动作-产生的刺激,对于那些无法用外部刺激与反应之间联结来说明的行为举止,提供了永远适用的现成的解释”。“动作-产生的刺激”现在已被许多心理学家用来作为分析“链索反应”过程的基础。

### 8.2.6 需要消减理论

赫尔(C L Hull)提出需要消减论。这种理论强调学习过程中包括四个要素:激励、提示、反应和报酬。激励是一种推动个体产生行动的内在刺激,有些激励属于生物性的,非学习的,如痛、渴、饥饿等;也有些是学习来的,如惧怕、社会需要等。激励是行为的基础,若无激励,个体没有行动,也就没有学习。

提示引发个体反应,并决定个体何时何地以及产生何种反应。提示的价值,在于它所具有的特征。提示可以是个体行动的目标,也可以是达到目标的启发,也可以兼具以上两种功能。

激励推动个体按提示反应，反应后获得报酬，则同样反应将会继续产生，若继续获得报酬，习惯即可形成。若反应后不予以报酬，反应重复的趋向将减低。因此激励因报酬而消减，反应因报酬而重现，这样构成学习。激励消减对个体来说就是需要的满足，需要的满足也正表示其需要程度的消减，所以这种理论称为需要消减论。

赫尔的体系表现在三本主要著作中：《行为的原理》《行为的基础》和《行为的体系》。在这一体系的最基本的形式中有十七项公式和十七项推论，这些形式作为交互作用的符号单位，阐明以下问题：①有机体带到学习情境中的非学习的刺激-反应联结和感觉能力；②能引起行为有效加强的动机过程和内驱力状态；③习惯形成的规律；④影响反应引出的那些非联想作用的因素；⑤反应倾向的反向条件抑制；⑥在一个以上刺激通过强化作用已和同一反应联系起来时使习惯力量和反应倾向两者都复杂化的因素；⑦当一个以上刺激在某一时刻同时呈现时，使刺激的兴奋特性复杂化的因素大于个体差异造成的学习公式中常量的变动。强化原理是这一体系的基石。在其初始形式中，赫尔的强化假说表明，一个反应由于一个基本需要或内驱力倾向得到满足而增强。

赫尔坚持学习的联结观点，但他与华生和格思里所不同的是，赫尔修改了 *S-R* 的公式，使之成为 *S-s-r-R* 的公式。其中 *S* 为外在环境刺激，*s* 为刺激痕迹，*r* 为运动神经冲动，*R* 为外部行为反应。赫尔认为，外在环境刺激消失后仍持续存在一段时间，成为刺激痕迹。该刺激痕迹导致了运动神经冲动，而该运动神经冲动最终导致了外部行为反应。赫尔希望借这样的修改能回避心理或意识的作用。此外，赫尔认为，引发行为的刺激极少是单个刺激，这也是行为难以预测的原因。通常是一组刺激及其相应的刺激痕迹共同作用，共同决定了行为，如图 8.2 所示。显然，赫尔想借此更加精确地说明和预测行为。

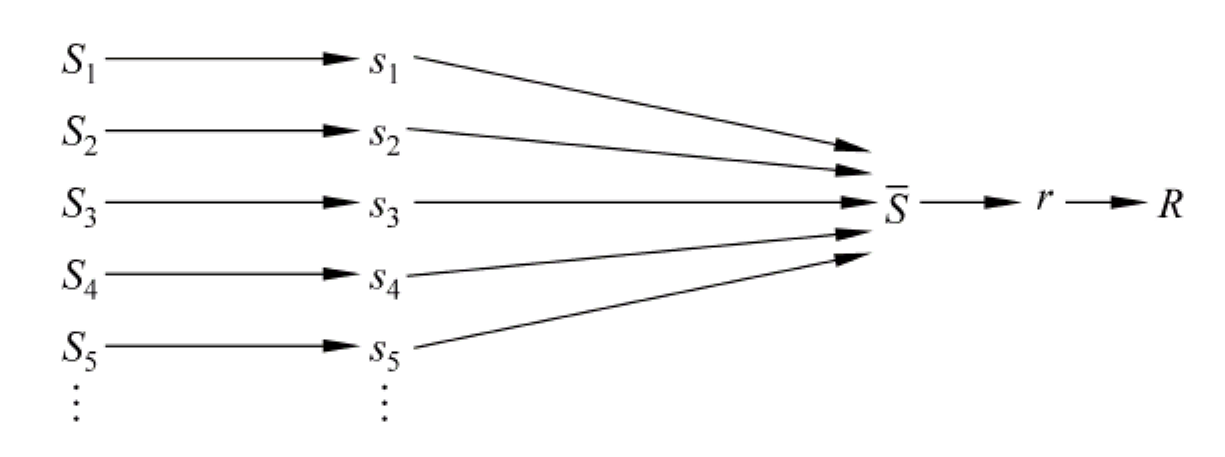



图 8.2　赫尔的多种刺激作用示意图

赫尔认为，刺激与反应在时间上的接近将加强该刺激引起该反应的可能性，在这一点上他接受了格思里的观点。但与格思里不同的是，接近并不是学习的唯一条件，甚至不是学习的充分条件，它只是一个必要条件。学习的另一个必要条件是强化，没有强化，便没有学习。强化是内驱力降低所必需的。赫尔将强化分为初级强化和次级强化两类。初级强化是内驱力降低的过程。满足这类强化的强化物主要是一些能满足有机体生物需要的物质，如食物、水等。另外一些中性刺激，本身并不具备降低内驱力的能量，但因为经常与初级强化物紧密相连，因而也获得了强化的效用，这就是次级强化。例如母亲每次喂奶时总是先将婴儿抱起来，喂奶是初级强化物，怀抱婴儿则成为次级强化物。

赫尔将刺激与反应的联结力量称为习惯强度，用 ${}_SH_R$ 表示。${}_SH_R$ 由接近和强化而得到加强。赫尔以公式表明了强化和习惯强度的函数关系。

$$ {}_SH_R = 1 - 10^{-0.0305n} \text{（}n\text{ 为强化的尝试次数）} \quad (8.1)$$

鉴于有机体的生物需要对于行为的激起作用，赫尔假设了内驱力（用 *D* 表示）这一中间

变量以替代生物需要的概念。赫尔认为,内驱力是一种有机体组织需要状态引起的刺激,是有机体一切行为反应的动力,它的力量可以由生物需要被剥夺时间的长短或所激起的行为的强度、力量等客观指标来加以确定。一个已经习得的反应是否发生的可能性(赫尔称之为反应势能,用${}_{S}E_{R}$ 表示),是受习惯强度和内驱力所决定的,用公式表示就是:

$$ {}_{S}E_{R} = D \cdot {}_{S}H_{R} \tag{8.2} $$

显然没有习惯强度就没有反应势能,而没有内驱力,反应也不可能发生。例如已经饱食的白鼠极少去按压杠杆,尽管它已经习得了按压杠杆的反应。

赫尔不仅认为每一种内驱力都有与之相联的特殊刺激,如干渴内驱力的特定刺激是嘴唇、喉头的干燥,而且认为内驱力也可以分为两种:原始内驱力和继起内驱力。原始内驱力与机体的生物需要相伴随,是维持生存所必需的,如饥、渴、性等。继起内驱力是指某一中性刺激由于曾伴随过原始内驱力所引起的反应,因而也就具有内驱力的性质,例如手被火炉烫伤,原始内驱力(解除烧伤之痛)驱使手立即缩回。与此相随的一个中性刺激,如看见火炉,可能由此而获得了强化的性质,即看见火炉,手即缩回。赫尔所说的这两种内驱力,实际上是无条件反射和条件反射的别名。

赫尔假定了最终定义为一种刺激的内驱力概念,来回避行为的动机作用问题,这反映了他的行为主义的客观立场。

在没有强化的条件下,一个不断重复的习得反应会逐渐削弱,以至消失,赫尔认为这是发生了抑制的缘故。抑制可分为两种:反应抑制(用 $I_R$ 表示),和条件抑制(用${}_{S}I_{R}$ 表示)。反应抑制是指有机体不断重复地活动产生了疲劳,而疲劳使动物削弱并最终终止其反应,当操作停止,反应抑制逐渐消失;如果在疲劳削弱其反应的同时,人为地匹配一个刺激,则该刺激也获得了抑制性力量,此为条件抑制。条件抑制是习得的,不会随时间而消失。既然这两种抑制都削弱反应,那么一个有效的反应势能(用${}_{S}E_{R}$ 表示)就应该是:

$$ {}_{S}E_{R} = D \cdot {}_{S}H_{R} - (I_{R} + {}_{S}I_{R}) \tag{8.3} $$

除此之外,赫尔还假定了抑制性潜能的存在,它能抵制习得性反应的发生,由于其作用发生不稳定,赫尔便以${}_{S}O_{R}$ 表示这种波动效应。所以,一个有效的反应势能还必须减去被动效应带来的抵制作用。即:

$$ {}_{S}E_{R} = [D \cdot {}_{S}H_{R} - (I_{R} + {}_{S}I_{R})] - {}_{S}O_{R} \tag{8.4} $$

由于抑制的增强,一个反应在未强化的情况下逐渐削弱以至消失,这即是行为的消退现象。在赫尔看来,行为的发生没有强化是不行的,但行为的消退则不是因为强化,而是抑制的影响。

赫尔在其学术生涯的后期对行为公式进行了修正,他认为强化是(用 $K$ 表示)和刺激强度(用 $V$ 表示)与习惯强度和内驱力一道共同决定了行为的发生,所以他的行为公式最终表述为:

$$ {}_{S}E_{R} = [D \cdot V \cdot K \cdot {}_{S}H_{R} - (I_{R} + {}_{S}I_{R})] - {}_{S}O_{R} \tag{8.5} $$

此外,赫尔还修正了他的内驱力降低起了强化作用的观点,认为不是内驱力本身的降低而是内驱力刺激的降低起了强化作用,如饮水直接消除了嘴唇、喉头的干燥,即刻降低了干渴内驱力,这一强化作用不必等到水经由胃的消化而至血液、至大脑的缓慢过程。

赫尔在其后期的修正中提出了一个非常重要的概念,即零星期待目标反应(以 $r_G$ 表示)。以白鼠走迷津为例,白鼠由起点走至终点获得食物,这种针对目标物的吃食反应即目

标反应。由于终点处的其他刺激经常与食物这一强化物紧密相连，因而成为次级强化物。依此类推，终点前面过道里的刺激，乃至通往终点的所有通道甚至起点处的刺激都逐渐具有了次级强化作用，成为次级强化物。它们使得白鼠在得到目标（即食物）以前对这些原本中性的刺激物产生了条件反应，由于这种反应是部分的、零碎的、于目标物之前的，故而称之为零星期待目标反应。

赫尔创造了零星期待目标概念，用以解释连锁反应的学习：动物在达到目的物之前的每一刺激既是对前一反应的强化，又是引发下一反应的刺激，由此而构成一个完整的行为链。赫尔希望借此概念的提出，能客观地说明人类行为的认知、意识过程，反映了新行为主义者在克服早期行为主义过于天真、简单等特点所作的努力。

赫尔花费了毕生的精力，以数学和逻辑方法为工具，提出了一个庞大的、无所不包的、企图说明“一切哺乳类动物行为”的理论体系，希望借此使心理学摆脱模糊、玄奥、不可测的旧轨，而成为一门完全客观的、像物理学和数学那样精确的自然科学，这是其他任何一位心理学家所不曾做过的。尽管最终的结果与初衷相去甚远，但这种尝试和努力本身是值得赞赏和令人鼓舞的，赫尔也以此对心理学作出了巨大贡献。

但是，赫尔的体系也遭受了众多的批评。他的体系过于庞大、复杂、精细，简直令人望而生畏，难以弄懂。赫尔的方法过于特殊而且缺少概括性，他的许多普遍的公式和参数往往来自于单一实验情境中的少量动物行为研究，而他又将它们用来概括和说明一般动物行为，这究竟有多大的代表性令人怀疑。而且赫尔假定和推断了许多内部状态变量，如刺激痕迹等等，但这些假定状态或变量又难以被证实。更一般地讲，赫尔的假设演绎系统是从经典几何学中借来的，应用于经验科学是否合适仍是一个问题。最后，人们还批评赫尔，说他是一个还原主义者，他以动物的刺激反应实验来解说人类的意识等高级认知活动，犯了还原论的错误。

在赫尔学术生涯的晚期以及在他逝世以后，他的许多同事和学生继承了他的事业，其中一部分沿着赫尔所开辟的动物学习研究领域和客观的方法路线继续工作，在坚持赫尔传统上表现得较为坚定、直接，这以肯尼依・斯彭斯等为代表；另外一些则坚持赫尔的基本立场和观点，但又不拘泥于此，他们将赫尔的概念和方法直接应用于人类行为中，如心理治疗、儿童发展，这主要以米勒、莫勒等为代表。

## 8.3　认知学习理论

与行为主义学习理论相对立，源自于格式塔学派的认知学习理论，经过一段时间的沉寂之后，再度复苏，从 20 世纪 50 年代中期之后，随着布鲁纳、奥苏伯尔等一批认知心理学家的大量创造性的工作，使学习理论的研究自桑代克之后又进入了一个辉煌时期，他们认为，学习就是面对当前的问题情境，在内心经过积极的组织，从而形成和发展认知结构的过程，强调刺激反应之间的联系是以意识为中介的，强调认知过程的重要性。因此，使认知学习论在学习理论的研究中开始占据主导地位。

认知是指认识的过程以及对认识过程的分析。美国心理学家吉尔伯特（G A Gilbert）认为：认知是一个人了解客观世界时所经历的几个过程的总称。它包括感知、领悟和推理等几个比较独特的过程，这个术语含有意识到的意思。认知的构造已成为现代教育心理学

家试图理解的学生心理的核心问题。认知学派认为学习在于内部认知的变化,学习是一个比 S-R 联结要复杂得多的过程。他们注重解释学习行为的中间过程,即目的、意义等,认为这些过程才是控制学习的可变因素。认知派学习理论的主要贡献如下。

(1) 重视人在学习活动中的主体价值,充分肯定了学习者的自觉能动性。

(2) 强调认知、意义理解、独立思考等意识活动在学习中的重要地位和作用。

(3) 重视了人在学习活动中的准备状态。即一个人学习的效果,不仅取决于外部刺激和个体的主观努力,还取决于一个人已有的知识水平、认知结构、非认知因素。准备是任何有意义学习赖以产生的前提。

(4) 重视强化的功能。认知学习理论由于把人的学习看成是一种积极主动的过程,因而很重视内在的动机与学习活动本身带来的内在强化的作用。

(5) 主张人的学习的创造性。布鲁纳提倡的发现学习论就强调学生学习的灵活性、主动性和发现性。它要求学生自己观察、探索和实验,发扬创造精神,独立思考,改组材料,自己发现知识、掌握原理原则,提倡一种探究性的学习方法。强调通过发现学习来使学生开发智慧潜力,调节和强化学习动机,牢固掌握知识并形成创新的本领。

认知学习理论的不足之处,是没有揭示学习过程的心理结构。我们认为学习心理是由学习过程中的心理结构,即智力因素与非智力因素两大部分组成的。智力因素是学习过程的心理基础,对学习起直接作用;非智力因素是学习过程的心理条件,对学习起间接作用。只有将智力因素与非智力因素紧密结合,才能使学习达到预期的目的。而认知学习理论对非智力因素的研究是不够重视的。

格式塔学派的学习理论、托尔曼的认知目的理论、皮亚杰的图式理论、维果斯基的内化论、布鲁纳的认知发现理论、奥苏伯尔的有意义学习理论、加涅的信息加工学习理论以及建构主义的学习理论均可作为认知派的代表性学说。认知主义学习理论的代表人物是皮亚杰、纽厄尔等。

### 8.3.1 格式塔学派的学习理论

格式塔学派又名完形学派,1912 年产生于德国,代表人物韦特海默、考夫卡、苛勒(K Kohler)。这一学派的学习理论是研究知觉问题时,针对桑代克的学习理论提出来的。他们强调经验和行为的整体性,反对行为主义的"刺激-反应"公式,于是他们重新设计了动物的学习实验。

苛勒从 1913—1917 年在一个岛上进行黑猩猩的学习实验。在一个典型的实验中,把黑猩猩关在笼中,笼外放有香蕉和一长一短的两只木杆。黑猩猩在笼内不能直接够到香蕉。黑猩猩用"手"够香蕉失败后,停止活动,四处张望,若有所思。之后,它突然起身,用短杆取得长杆,再用长杆够到了香蕉。这一系列动作是一气呵成的。由此,苛勒认为,黑猩猩对问题的解决是由于突然领悟即顿悟而实现的,学习不是逐渐地试误过程,而是对知觉经验的重新组织,是对情境关系的顿悟。

格式塔学派的基本观点如下。

(1) 学习是组织一种完形。

完形派认为,学习是组织一种完形。完形或称"格式塔"指的是事物的式样和关系的认知。学习过程中问题的解决,是由于对情境中事物关系的理解而构成一种完形来实现的。

学习即黑猩猩在实验情境中发现关系(木杆是获得香蕉的工具),从而弥和缺口,构成完形。完形派认为,无论是运动的学习、感觉的学习、感觉运动的学习和观念的学习,都在于发生一种完形的组织,并非各部分间的联结。

(2) 学习是通过顿悟实现的。

完形派认为学习的成功和实现完全是“顿悟”的结果,即突然地理解了,而不是“试误”,“尝试与错误”。顿悟是对情境全局的知觉,是对问题情境中事物关系的理解,也就是完形的组织过程。

完形派用来证明学习过程是领悟而非试误的主要证据:①从不能到能之间突然转变;②学到的东西能良好的保持,而不是重复出现错误。他们指出,由于桑代克所设置的问题情境不明确,从而导致了盲目的尝试错误学习。

对完形派学习理论的评价如下。

(1) 完形派学习理论具有辩证的合理因素,主要表现在它肯定了意识的能动作用,强调了认知因素(完形的组织)在学习中的作用。由此弥补了桑代克学习理论的缺陷,认为刺激与反应之间的关系是间接的,不是直接的,是以意识为中介的。完形派对试误说的批判,也促进了学习理论的发展。

(2) 完形派在肯定顿悟的同时,否定试误的作用,是片面的。试误与顿悟是学习过程的不同阶段,或不同的学习类型。试误往往是顿悟的前奏,顿悟又往往是试误的必然结果,二者不是相互排斥、对立的,而应是相互补充的。完形派的学习理论不够完整,也不够系统,其影响在当时远不及桑代克的联结说。

## 8.3.2 认知目的理论

托尔曼(E C Tolman)认为自己是一名行为主义者。他对各派采取兼容并包的态度,以博采众家之长而著称。他既欣赏联结派的客观性和测量行为方法的简便,又受到格式塔整体学习观的影响。他的学习理论有很多名称,如符号学习说、学习目的说、潜伏学习说、期待学习说。他坚持主张理论要用完全客观的方法检验。然而许多人认为他是研究动物学习行为最有影响的认知主义者。受格式塔学派的影响,他强调行为的整体性。他认为整体行为是指向一定目的的,而有机体对环境的认知是达到目的的手段。他不同意把情境(刺激)与反应之间看成是直接的联系,即 *S-R*。他提出“中介变量”的概念,认为中介变量是介于实验变量和行为变量之间并把二者联系起来的因素。具体说,中介变量就是心理过程,由心理过程把刺激与反应联结起来。因此 *S-R* 的公式应 *S-O-R*,*O* 即代表中介变量。他的学习理论就是从上述观点出发,通过对动物学习行为全过程的考察而提出的。

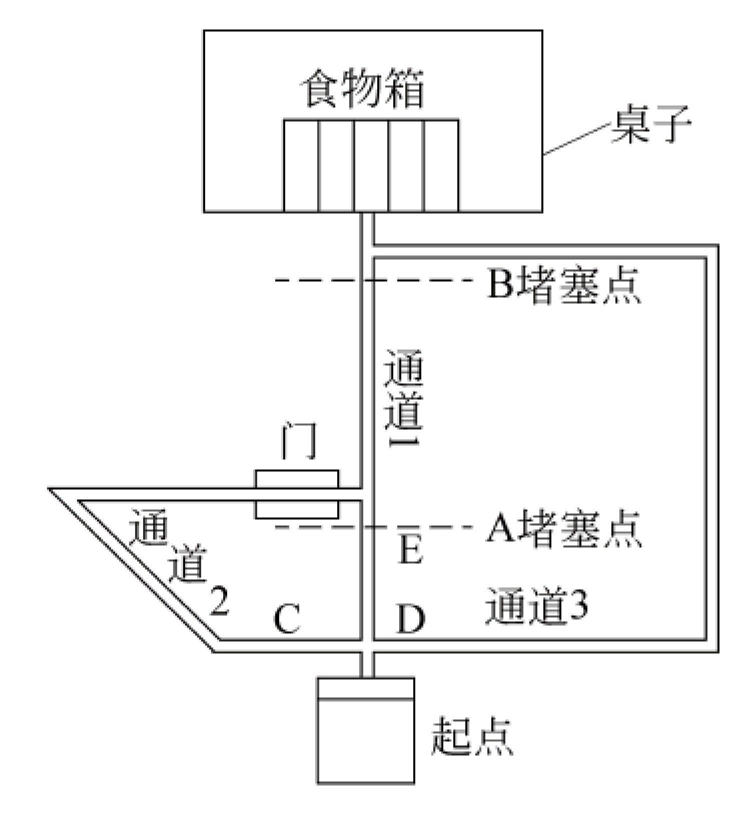


图 8.3 白鼠迷津实验

托尔曼于 1930 年设计并进行了白鼠高架迷津方位实验。在这种迷津中设置了白鼠通向食物箱的长短不等的三条通道(见图 8.3)。

首先让白鼠在迷津内经过探索,熟悉这三条通道,然后将白鼠放进起点箱内,观察它们的行为。结果发现,白鼠首先选择通向食物距离最短的通道 1,当通道在 A 处堵

塞时,它们便在通道2和通道3中选择了较短的通道2;而通道2必经的B处也被堵塞时,它们才不得不选择较漫长的通道3。

托尔曼认知目的理论的基本观点如下。

(1) 学习是有目的的。

托尔曼认为动物学习是有目的的,其目的就是获得食物。他不同意桑代克等人认为学习是盲目的观点。动物在迷津中的试误行为是受目标指引的,是指向食物的,是不达目的不罢休的。他认为学习就是期望的获得。期望是个体关于目标的观念。个体通过对当前的刺激情境的观察和已有的过去经验而建立起对目标的期望。

(2) 对环境条件的认知是达到目的的手段或途径。

托尔曼认为有机体在达到目的的过程中,会遇到各式各样的环境条件,他必须认知条件,才能克服困难,达到目的。所以,对环境条件的认知是达到目的的手段或途径。(托尔曼用"符号"代表有机体对环境条件的认知。)学习不是简单地、机械地形成运动反映,而是学习达到目的的符号,形成"认知地图"。所谓认知地图是动物在头脑中形成的对环境的综合表象,包括路线、方向、距离,甚至时间关系等信息。这是个较模糊的概念。

总之,目的和认知是托尔曼学习理论中的两个重要中介变量,所以称他的学习理论为认知目的理论。

托尔曼认知目的理论中重视行为的整体性、目的性,提出中介变量的概念,重视在刺激与反应之间的心理过程,强调认知、目的、期望等在学习中的作用,是进步,应给予肯定。托尔曼理论中的一些术语,如"认知地图"没有明确的界定;对人类的学习与动物的学习也没有从本质上进行区分,因而是机械注意的,这使得他的理论不能成为一个完整的合理的体系。

### 8.3.3 认知发现理论

布鲁纳是美国当代著名的认知心理系家。1960年,他同乔治·米勒一起创建了哈佛大学认知研究中心。他是美国认知学说的主要代表人物。

布鲁纳的认知学习理论受完形说、托尔曼的思想和皮亚杰发生认识论思想的影响,认为学习是一个认知过程,是学习者主动地形成认知结构的过程。而布鲁纳的认知学习理论与完形说及托尔曼的理论又是有区别的。其中最大的区别在于完形说及托尔曼的学习理论是建立在对动物学习进行研究的基础上的,所谈的认知是知觉水平上的认知,而布鲁纳的认知学习理论是建立在对人类学习进行研究的基础上的,所谈认知是抽象思维水平上的认知。其基本观点主要表现在以下三个方面。

#### 1. 学习是主动地形成认知结构的过程

认知结构是指一种反映事物之间稳定联系或关系的内部认识系统,或者说,是某一学习者的观念的全部内容与组织。人的认识活动按照一定的顺序形成,发展成对事物结构的认识后,就形成了认知结构,这个认知结构就是类目及其编码系统。布鲁纳认为,人是主动参加获得知识的过程的,是主动对进入感官的信息进行选择、转换、存储和应用的。也就是说人是积极主动地选择知识的,是记住知识和改造知识的学习者,而不是一个知识的被动的接受者。布鲁纳认为,学习是在原有认知结构的基础上产生的,不管采取的形式怎样,个人的

学习,都是通过把新得到的信息和原有的认知结构联系起来,去积极地构建新的认知结构的。

布鲁纳认为学习包括着三种几乎同时发生的过程,这三种过程是:新知识的获得、知识的转化、知识的评价。这三个过程实际上就是学习者主动地建构新认知结构的过程。

### 2. 强调对学科的基本结构的学习

布鲁纳非常重视课程的设置和教材建设,他认为,无论教师选教什么学科,务必要使学生理解学科的基本结构,即概括化了的基本原理或思想,也就是要求学生以有意义地联系起来的方式去理解事物的结构。布鲁纳之所以重视学科的基本结构的学习,是受他的认知观和知识观的影响的。他认为,所有的知识,都是一种具有层次的结构,这种具有层次结构性的知识可以通过一个人发展的编码体系或结构体系(认知结构)而表现出来。人脑的认知结构与教材的基本结构相结合会产生强大的学习效益。如果把一门学科的基本原理弄通了,则有关这门学科的特殊课题也不难理解了。

在教学当中,教师的任务就是为学生提供最好的编码系统,以保证这些学习材料具有最大的概括性。布鲁纳认为,教师不可能给学生讲遍每个事物,要使教学真正达到目的,教师就必须使学生能在某种程度上获得一套概括了的基本思想或原理。这些基本思想、原理,对学生来说,就构成了一种最佳的知识结构。知识的概括水平越高,知识就越容易被理解和迁移。

### 3. 通过主动发现形成认知结构

布鲁纳认为,教学一方面要考虑人的已有知识结构、教材的结构,另一方面要重视人的主动性和学习的内在动机。他认为,学习的最好动机是对所学材料的兴趣,而不是奖励竞争之类的外在刺激。因此,他提倡发现学习法,以便使学生更有兴趣、更有自信地主动学习。

发现法的特点是关心学习过程胜于关心学习结果。具体知识、原理、规律等让学习者自己去探索、去发现,这样学生便积极主动地参加到学习过程中去,通过独立思考,改组教材。"学习中的发现确实影响着学生,使之成为一个'构造主义者'"。学习是认知结构的组织与重新组织。它既强调已有知识经验的作用,也强调学习材料本身的内在逻辑结构。

布鲁纳认为发现学习的作用有以下几点。

(1) 提高智慧的潜力。

(2) 使外来动因变成内在动机。

(3) 学会发现。

(4) 有助于对所学材料保持记忆。

所以,认知发现说是值得特别重视的一种学习理论。认知发现说强调学习的主动性,强调已有认知结构、学习内容的结构、学生独立思考等的重要作用。这些对培育现代化人才是有积极意义的。

## 8.3.4 认知同化理论

奥苏伯尔是美国纽约州大学研究院的教育心理学教授,是认知派的代表人物之一。他从 21 世纪 50 年代中期开始致力于有意义言语材料的学习与保持的研究。他的理论在

60 年代提出,受到中小学教师的欢迎。1976 年获得美国心理学会的“桑代克奖”。

奥苏伯尔从学生学习的方式上将学习分为接受学习与发现学习,从学习内容与学习者认知结构的关系上又将学习分为有意义学习和机械学习。

奥苏伯尔认为学生的学习主要是接受学习,而不是发现学习。接受学习是教师将学习内容以定论的形式直接呈现给学生,教师传授,学生接受。自实行班级授课制以来,接受学习一直是课堂学习的主要形式,但这种学习形式一直被误解为机械学习。奥苏伯尔认为接受学习既可以是有意义的,也可以是机械的,只是因为一些教师使学生进行的是机械学习而采取的又是接受学习方式,才使接受学习被认为是机械的。同样,发现学习既可以是有意义的学习,也可以是机械学习,那种只发现点滴的事实,而不理解其中的规律的发现学习便是机械的发现学习。奥苏伯尔认为学校中的学习应该是有意义的接受学习和有意义的发现学习,但他更强调有意义的接受学习,认为它可以在短时间内使学生获得大量的系统知识,这正是教学的首要目标。

奥苏伯尔认为,有意义学习的实质就是以符号代表的新观念与学习者认知结构中原有的适当观念建立起非人为的和实质性联系的过程。奥苏伯尔所定义的认知结构是一个人的观念的全部内容与组织或一个人在某个知识领域的观念的内容与组织。认知结构中原有的知识是观念的支架,或称之为起固定作用的观念。有意义学习的过程就是新观念被认知结构中起固定作用的观念同化、储存并相互作用,原有的观念同时发生变化,新知识纳入原有的认知结构中,从而获得意义。

所谓非人为的、实质性的联系是指新知识与学习者认知结构中已有的表象,已有意义符号、概念或命题的联系。建立起非人为的、实质性的联系是有意义学习的两个标准。非人为的联系是指新的观念与原有观念建立了内在的联系,而不是任意的联系;实质性是指用不同语言或其他符号表达的同一认知内容的联系。机械学习中的新知识与原有认知结构只能建立起人为的、表面的、非实质性的联系。

奥苏伯尔认为有意义学习必须具备以下三个前提条件。

(1) 学习材料本身必须具备逻辑意义。所谓逻辑意义是指学习材料可以和学习者认知结构中的适当观念建立起非人为的和实质性的联系。

(2) 学习者必须具备有意义学习的倾向,即积极主动地把新知识与学习者认知结构中原有的适当知识联系起来的倾向性。

(3) 学习者认知结构中必须具有同化新知识的适当观念。

以上三个条件必须同时具备,才能实现有意义学习。学习者必须积极主动地使具有逻辑意义的新知识与其原有认知结构中的有关的旧知识发生相互作用,旧知识得到改造,新知识获得实际意义。奥苏伯尔认为在有意义学习中,影响新知识的学习的最重要的条件是学习者。

有意义学习的过程即原有观念对新观念加以同化的过程。奥苏伯尔称自己的学习理论为同化论。原有观念一般通过三种方式对新观念进行同化,即类属学习、总括学习、并列结合学习。类属学习是把新观念归入认知结构中原有观念的适当部分,并使之相互联系的过程。

类属学习过程中,原有观念是总观念,新学习的观念是从属观念,因而这种学习称为下位学习。类属学习又有两种形式:派生类属学习和相关类属学习。派生类属学习中,新观

念只是使原有总观念扩张，并不能使原有观念的本质发生改变。

相关类属学习中认知结构中原有观念是一个总的观念，所学的新观念只是原有观念的加深、修饰或限定，通过同化，总观念的本质要发生变化。

总括学习是指在若干已有的从属观念的基础上归纳出一个总观念。例如，掌握了铅笔($a_1$)、橡皮($a_2$)、笔记本($a_3$)等观念之后，再学习更高一级的总观念“文具”($A$)时，原有的从属观念 $a_n$ 可以为学习总观念($A$)服务。总括学习所形成的新观念在概括和包容程度上高于原有的一些观念，所以称为上位学习。

并列结合学习是指导新观念与认知结构中的原有观念，既非从属关系，也非总括关系，只是和原有认知结构中的整个内容具有一般的联系。例如，新学习的观念是“质量与能量的关系”($A$)，原有观念是“热和体积”($B$)、“遗传和变异”($C$)、“需求与价格”($D$)之间的关系。在此条件下，新观念既不能类属于某一特殊关系，也不能总括原有的关系，但它们具有某种共同的关键属性。由于新知识与原有知识具有某种共同属性，因而也可以被原有的知识同化，获得意义。

奥苏伯尔假定：认知结构本身是一个按层次组织的观念系统，最概括的观念处于这个结构的顶端，在其下面是概括性较少的观念，较分化的从属观念和具体材料。在类属学习中，学习者利用认知结构中抽象概括水平较高的观念固定和同化新观念，而每一次类属都进一步证实、扩展或深化了原来的“固定观念”，使原来的认知结构不断分化。奥苏伯尔非常强调类属学习，认为多数有意义学习都具有自上而下渐进分化的特征。总括学习通常在进行归纳、推理或综合部分与整体的关系时需要；而在并列结合学习中，由于缺乏最适当的起固定作用的观念，学习时一般比较困难，而且不易保持。

### 8.3.5 信息加工学习理论

加涅(R M Gagne)是美国佛罗里达州大学的教育心理学教授。他的学习理论是在行为主义和认知观点相结合的基础上，在 20 世纪 70 年代之后，运用现代信息论的观点和方法，通过大量实验研究工作建立起来的。他认为学习过程是信息的接收和使用过程。学习是主体和环境相互作用的结果，学习者内部状况与外部条件是相互依存、不可分割的统一体。

加涅认为，学习是学习者神经系统中发生的各种过程的复合。学习不是刺激反应间的一种简单联结，因为刺激是由人的中枢神经系统以一些完全不同的方式来加工的，了解学习也就在于指出这些不同的加工过程是如何起作用的。在加涅的信息加工学习论中，学习的发生同样可以表现为刺激与反应，刺激是作用于学习者感官的事件，而反应则是由感觉输入及其后继的各种转换而引发的行动，反应可以通过操作水平变化的方式加以描述。但刺激与反应之间，存在着“学习者”“记忆”等学习的基本要素。学习者是一个活生生的人，他们拥有感官，通过感官接受刺激；他们拥有大脑，通过大脑以各种复杂的方式转换来自感官的信息；他们有肌肉，通过肌肉动作显示已学到的内容。学习者不断接受到各种刺激，被组织进各种不同形式的神经活动中，其中有些被储存在记忆中，在作出各种反应时，这些记忆中的内容也可以直接转换成外显的行动。加涅将学习过程看作是信息加工流程。1974 年，他描绘出一个典型的学习结构模式图(见图 8.4)。

加涅的学习结构模式分两个部分。第一部分是右边的结构叫操作记忆，是一个信息流。来自环境的刺激作用于学习者的感受器，然后到达感觉记录器，信息在这里经过初步的选择

处理,停留的时间还不到一秒钟,便进入短时记忆,信息在这里也只停留几秒钟,然后进入长时记忆。以后当需要回忆时,信息从长时记忆中提取而回到短时记忆中,然后到达反应发生器,信息在这里经过加工便转化为行为,作用于环境,这样就发生了学习。第二部分是左边的结构,包括预期事项(期望)和执行控制两个环节。预期环节起着定向的作用,使学习活动沿着一定方向进行。执行环节起调节、控制作用,使学习活动得以实现。第二部分的功能是使学习者引起学习、改变学习,加强学习和促进学习,同时使信息流激化、削弱或改变方向。

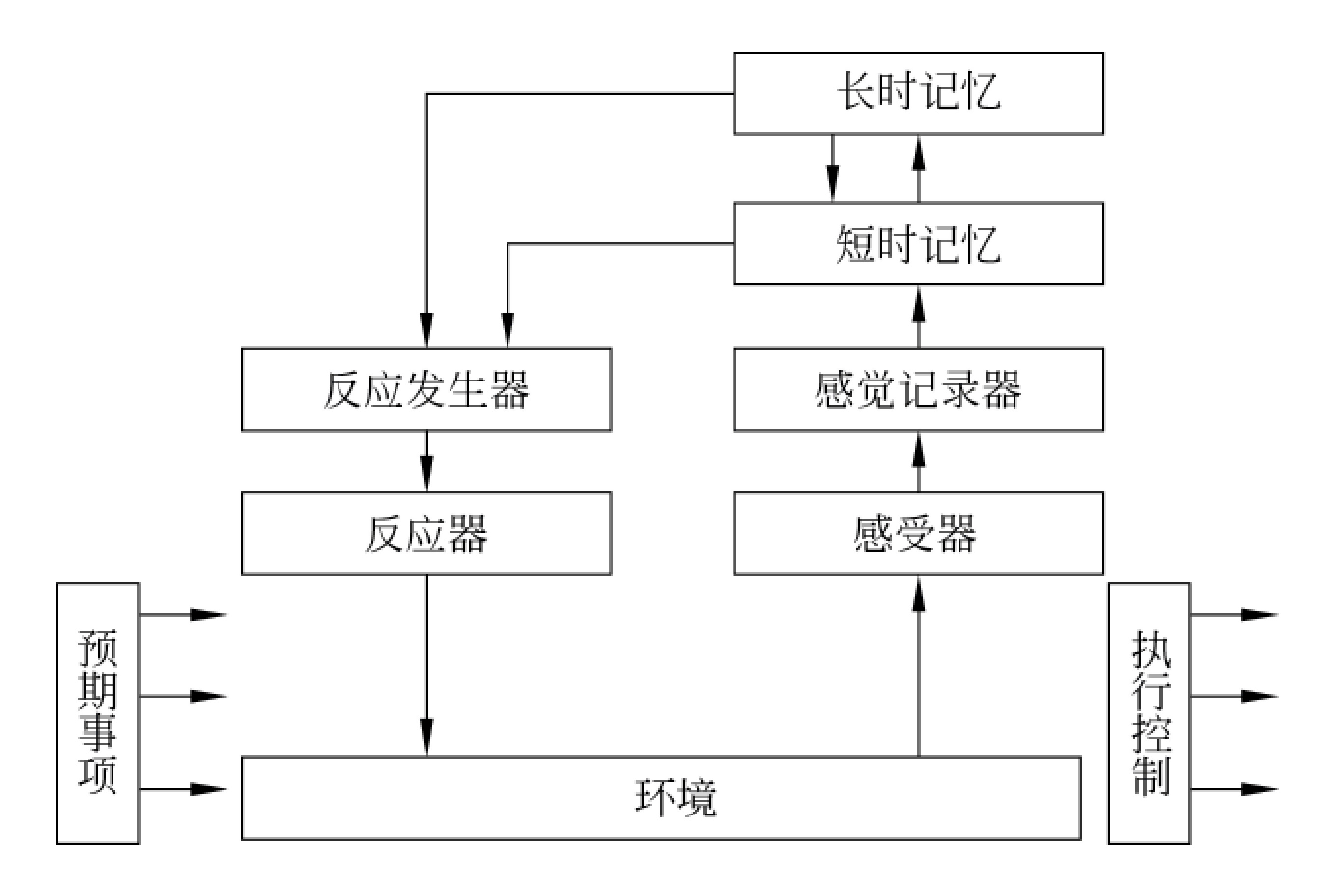



图 8.4　学习结构模式图

加涅根据信息加工理论提出了学习过程的基本模式,认为学习过程就是一个信息加工的过程,即学习者对来自环境刺激的信息进行内在的认知加工的过程,并具体描述了典型的信息加工模式。认为学习可以区别出外部条件和内部条件,学习过程实际上就是学习者头脑中的内部活动,与此相应,把学习过程划分为八个阶段:动机阶段、了解阶段、获得阶段、保持阶段、回忆阶段、概括阶段、作业阶段和反馈阶段(见图 8.5)。

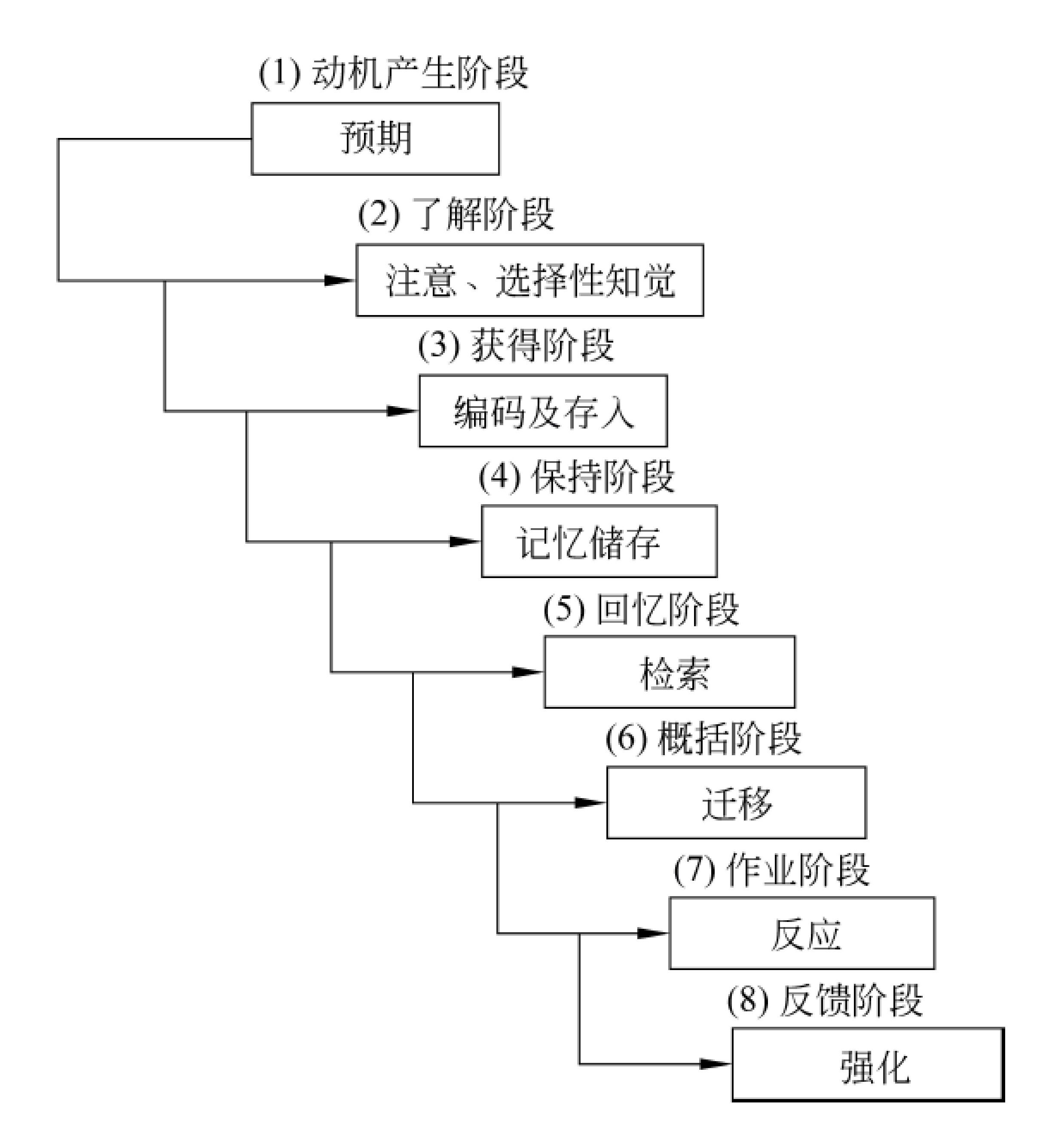



图 8.5　加涅的八个学习阶段及其相应的心理过程图

(1) 动机产生阶段，与之相应的心理过程是期望。学习要先有动机，动机可以与学习者的期望建立联系。期望是目标达到时所能得到的报酬、结果或奖励，是完成任务的动力，能给学习者指明方向和道路。

(2) 了解阶段，与之相应的心理过程是注意、选择性知觉。加涅认为注意是一个短暂的内部状态，对学习有定势作用，也起着执行控制作用。教学要引起学生这种注意，通过口头指导语把注意引向学习有关的某一方面，可使学生有选择地感知其所处情况中的某些刺激。

(3) 获得阶段，与之相应的心理过程是编码、存入。在这一阶段，所学知识到达短时记忆，并转入长时记忆。编码就是对获得的信息进行加工整理，以便和原有信息相联系并形成系统，存入长时记忆。

(4) 保持阶段，与之相应的心理过程是记忆储存。知识到达长时记忆后，还要对材料继续加工，使之能永久保持。

(5) 回忆阶段，与之相应的心理过程是检索。回忆是指能将所学材料准确地重现出来，是通过检索实现的。检索是在外部刺激作用下，按一定方向进行的寻找过程。

(6) 概括阶段，与之相应的心理过程是迁移。对学习材料进行总结、整理、归纳，形成体系或结构，并能将知识和技能应用到各种新的情境中，其实质为学习迁移。

(7) 作业阶段，与之相应的心理过程是反应。是学习者将学习付诸行动，通过新作业和新操作的完成，表现出学习者学到了什么。

(8) 反馈阶段，与之相应的心理过程是强化的。在这一阶段，学习者完成了新作业并意识到自己已达到预期目标，从而使第一阶段所建立的预期和动机，在最后阶段得到证实和强化。加涅认为，强化主宰着人类的学习。

加涅认为新的学习一定要适合学习者当时的认知发展水平，即学习者已经发展形成认知结构。认为学习要在学习者内在认知结构和新输入的信息之间，建立起相互联系和相互配合的新结构。学习的理想条件是要把新输入的信息与学习者已有认知结构之间所存在的矛盾或差距，给以适当调整。这样，新信息能纳入已有认知结构中去，并建立新的认知结构。新的认知结构又作为高一级学习的基础，这样认知结构得到逐级发展和提高。

所谓指导结果指教师要给学生以最充分地指导，使学生沿着仔细规定的学习程序，引导学生一步一步地循序渐进地进行学习。指导法是依据他对教学目标和能量的理解而提出来的。加涅认为教学的主要目标是发展能量(即能力)，而发展能量的关键在于掌握大量有组织的知识是一个金字塔型的知识系统。教学目标确定之后，教师首先应进行任务分析，任务分析是自上而下进行的。为使学生获得终极行为，学生需要学会做哪些事？必须表现出什么起点行为？这样就构成了层次学习图。

加涅的学习理论注重学习的内部条件和学习的层次，重视系统知识的系统教学及教师循序渐进的指导作用，为控制教学提供了一定的依据。他的理论直接涉及课堂教学，因而对实际教学都有积极的意义和一定的参考价值。加涅运用信息论、控制论的观点和方法对学习问题进行有意义的探索。他试图兼收行为主义和认知派学习理论中的一些观点来建立自己的学习理论，反映了西方学习理论发展的一种趋势。他的学习理论，把能力(他所说的能量)仅仅归结为大量有组织的知识，具有一定的片面性，忽视了思维和智力技能的作用及其培养。

### 8.3.6 建构主义的学习理论

建构主义(Constructivism)是学习理论中行为主义发展到认知主义以后的进一步发展,即向着与客观主义更为对立的另一方面发展。建构主义的核心观点认为:第一,认识并非主体对于客观实在的简单的、被动的反映(镜面式反映),而是一个主动的建构过程,即所有的知识都是建构出来的;第二,在建构的过程中主体已有的认知结构发挥了特别重要的作用,而主体的认知结构亦处在不断的发展之中。皮亚杰和维果斯基是寻构主义的先驱者。尽管皮亚杰高度强调每个个体的新创造;而维果斯基更关心知识的工具即文化和语言的传递,但在基本方向上,皮亚杰和维果斯基都是建构主义者。

现代的建构主义又可以区分为极端建构主义和社会建构主义。极端建构主义有两个基本特征:首先是突出强调认识活动的建构性质,认为一切知识都是主体的建构,我们不可能具有对外部世界的直接认识,认识活动就是一个“意义赋予”(Sense Making)的过程,即是主体依据自身已有的知识和经验建构出对外部世界的意义;其次是对认识活动的“个体性质”的绝对肯定,认为各个主体必然地具有不同的知识背景和经验基础(或不同的认知结构),因此,即使就同一个对象的认识而言,相应的认识活动也不可能完全一致,而必然地具在个体的特殊性。在极端建构主义者看来,个人的建构有其充分的自主性,即是一种高度自主的活动,也就是说“一百个人就是一百个主体,并会有一百个不同的建构”。也正是在这样的意义上,极端建构主义也常常被称作“个人建构主义”。社会建构主义的核心在于对认识活动的社会性质的明确肯定,认为社会环境、社会共同体对于主体的认识活动有重要作用,个体的认识活动是在一定的社会环境中得以实现的,所谓的“意义赋予”包含有“文化继承”的意义,即经由个体的建构活动所产生的“个体意义”事实上包含了对于相应的“社会文化意义”的理解和继承。

建构主义认为学习是学习者运用自己的经验去积极地建构对自己富有意义的理解,而不是去理解那些用已经组织好的形式传递给他们的知识。学习者对外部世界的理解是他或她自己积极建构的结果,而不是被动地接受别人呈现给他们的东西。建构主义者认为知识是个体对现实世界建构的结果。根据这种观点,学习发生于对规则和假设的不断创造,以解释所观察到的现象。而当学习者对现实世界的原有观念与新的观察之间出现不一致,原有观念失去平衡时,便产生了创造新的规则和假设的需要。可见,学习活动是一个创造性的理解过程。相对于一般的认识活动而言,学习主要是一个“顺应”的过程,即认知结构的不断变革或重组,而认知结构的变革或重组又正是新的学习活动与认知结构相互作用的直接结果。按照建构主义的观点,“顺应”或认知结构的变革或重组正是主体主动的建构活动。建构主义强调学习者的积极主动性、强调新知识与学习者原有知识的联系、强调将知识应用于真实的情境中而获得理解。美国心理学家维特罗克(M C Wittrock)提出的学生学习的生成过程模式较好地说明了学习的这种建构过程。维特罗克认为学习的生成过程是学习者原有的认知结构即已经储存在长时记忆中的事和脑的信息加工策略,与从环境中接受的感觉信息(新知识)相互作用,主动地选择信息和注意信息,以及主动地建构信息的意义的过程。

学生的学习是在学校这样一个特定的环境中,是在教师的直接指导下进行的,主要是一种文化继承的行为,即学习这一特殊的建构活动具有明显的社会性质,是一种高度了的社会行为。学习并非一种孤立的个人行为,适当的环境不仅是学习的一个必要条件,而且也在很

大程度上决定了智力的发展方向。

根据建构主义的基本立场，教师和学生与学生和学生之间的相互作用对学习活动有重要影响。小组合作学习近年来受到普遍的重视，因为它为更充分地去实现“社会相互作用”提供了现实的可能性。正是基于这样的认识，人们提出了“学习共同体”的概念，即认为学习活动是由教师和学生所组成的共同体共同完成的。也就是说，学习不能被看作是孤立的个人行为，而是“学习共同体”的共同行为，或者说共同行为与个人行为之间存在着一种相互依赖、相互促进的辩证关系。此外，我们还应看到整体性的社会环境和文化传统对于个人的学习活动亦有十分重要的影响。

传统的认知派学习理论认为，学习的结果是形成认知结构，这是高度结构化的知识，是按概括水平的高低分层次排列的。

建构主义认为学生学习的结果是建构围绕着关键概念的网络结构知识，包括事实、概念、概括化以及有关的价值、意向、过程知识、条件知识等等。其中关键概念是结构性知识，而网络的其他方面含有非结构性知识。因此，建构主义学习理论认为学习的结果既包括结构性知识，也包括非结构性知识，而且认为这是高级学习的结果。

斯皮罗(Spiro)等人认为学习可以分为初级学习和高级学习。初级学习是学习的低级阶段，在该阶段，学生知道一些重要的概念和事实，在测验中能将所学的东西按原样再生出来，这里所涉及的内容主要是结构良好的领域(Well-Structured Domains)。高级学习要求学生把握概念的复杂性，并广泛而灵活地运用到具体情境中，这时所涉及的是大量结构不良领域(Ill-Structured Domains)的问题。概念的复杂性和概念实例间的差异性是结构不良领域的两个主要特点。斯皮罗认为结构不良领域是普遍存在的，只要将知识运用到具体情境中去，都有大量的结构不良的特征。因此，在解决实际问题时，往往不能靠简单地提取出某一个概念原理，而是要通过多个概念原理以及大量的经验背景的共同作用而实现。

建构主义学习理论是学习理论的一种新的发展。该理论强调学习过程中的积极主动性、对新知识的意义的建构性和创造性的理解，强调学习是社会性质，重视师生之间和学生与学生之间的社会相互作用对学习的影响，将学习分为初级学习和高级学习，强调学生通过高级学习建构网络结构知识，并在教学目标、教师的作用、促进教学的条件以及教学方法和设计等方面提出了一系列新颖而富有创见的主张，这些观点和主张对于进一步认识学习的本质，揭示学习的规律，深化教学改革都具有积极意义。

建构主义学习理论是在吸收了各种学习理论观点基础上形成和发展起来的，其中一些观点的论述往往失之偏颇，甚至相互对立，这在一定程度上暴露了该理论的不足之处，有待于进一步发展和完善。

## 8.4　人本学习理论

人本主义心理学是 20 世纪 50～60 年代在美国兴起的一种心理学思潮，其主要代表人物是马斯洛(A Maslow)和罗杰斯(C R Rogers)。人本主义的学习与教学观深刻地影响了世界范围内的教育改革，是与程序教学运动、学科结构运动齐名的 20 世纪三大教学运动之一。

人本主义心理学家认为，要理解人的行为，就必须理解行为者所感知的世界，即要知道

从行为者的角度来看待事物。在了解人的行为时,重要的不是外部事实,而是事实对行为者的意义。如果要改变一个人的行为,首先必须改变他的信念和知觉。当他看问题的方式不同时,他的行为也就不同了。换言之,人本主义心理学家试图从行为者,而不是从观察者的角度来解释和理解行为。下面我们介绍人本主义学习理论代表人物——罗杰斯的学习理论。

罗杰斯认为,可以把学习分成两类。一类学习类似于心理学上的无意义音节的学习。罗杰斯认为这类学习只涉及心智,是一种"在颈部以上"发生的学习。它不涉及感情或个人意义,与完整的人无关。另一类是意义学习。所谓意义学习,不是指那种仅仅涉及事实累积的学习,而是指一种使个体的行为、态度,个性以及在未来选择行动方针时发生重大变化的学习。这不仅仅是一种增长知识的学习,而且是一种与每个人各部分经验都融合在一起的学习。

罗杰斯认为,意义学习主要包括如下四个要素。

(1) 学习具有个人参与(Personal Involvement)的性质,即整个人(包括情感和认知两方面)都投入学习活动。

(2) 学习是自动自发的(Self-initiated),即便在推动力或刺激来自外界时,要求发现、获得、掌握和领会的感觉也是来自内部的。

(3) 全面发展,也就是说,它会使学生的行为、态度、人格等获得全面发展。

(4) 学习是由学生自我评价的(Evaluated by the Learner),因为学生最清楚这种学习是否满足自己的需要、是否有助于获得他想要知道的东西、是否明了自己原来不甚清楚的某些方面。

罗杰斯认为,促进学生学习的关键不在于教师的教学技巧、专业知识、课程计划、视听辅导材料、演示和讲解、丰富的书籍等等,而在于教师和学生之间特定的心理气氛因素。那么,好的心理气氛因素包括什么呢?罗杰斯给出了自己的解释:①真实或真诚:教师作为学习的促进者,表现真我、没有任何矫饰、虚伪和防御;②尊重、关注和接纳:教师尊重学习者的意见和情感,关心学习者的方方面面,接纳作为一个个体的学习者的价值观念和情感表现;③移情性理解:教师能了解学习者的内在反应,了解学生的学习过程。在这种心理气氛下进行的学习,是以学生为中心的,教师是学习的促进者、协作者或者说是伙伴,朋友、学生才是学习的关键,学习的过程就是学习的目的所在。

总之,罗杰斯等人本主义心理学家从他们的自然人性论、自我实现论出发,在教育实际中倡导以学生经验为中心的"有意义的自由学习",对传统的教育理论造成了冲击,推动了教育改革运动的发展。这种冲击和促进表现在:突出情感在教学中的地位和作用,形成了一种以情感作为教学活动的基本动力的新的教学模式;以学生的"自我"完善为核心,强调人际关系在教学过程中的重要性;把教学活动的重心从教师引向学生,把学生的思想、情感、体验和行为看作是教学的主体,从而促进了个别化教学的发展。

可以看到,人本主义学习理论中的许多观点都是值得我们借鉴的。比如,教师要尊重学生、真诚地对待学生;让学生感到学习的乐趣,自动自发地积极参与到教学中;教师要了解学习者的内在反应,了解学生的学习过程;教师要作为学习的促进者、协作者或者说是学生的伙伴、朋友,等等。但是,我们也需要看到,罗杰斯过分否定教师的作用,这是不太正确的。在教学中,我们既要强调学生的主体地位,也不能忽视教师的主导作用。

# 8.5　观察学习理论

班图拉对心理学的杰出贡献在于他发掘了前人所忽视的学习形式——观察学习，给予观察学习以应有的重视和地位。他提出的观察学习模式同经典条件反射和操作条件反射一起被称为解释学习的三大工具。观察学习理论有时也称为社会学习理论。班图拉的学习理论不回避人的行为的内部原因，相反，它重视符号、替代、自我调节所起的作用。因此，班图拉的社会学习论被称为认知行为主义。

班图拉在其观察学习的研究中，注重社会因素的影响，改变了传统学习理论重个体轻社会的思想倾向，把学习心理学的研究同社会心理学的研究结合在一起，对学习理论的发展做出了独树一帜的贡献。班图拉吸收认知心理学的研究成果，把强化理论与信息加工理论有机地结合起来，改变了传统行为主义重刺激-反应和轻中枢过程的思想倾向，使解释人的行为的理论参照点发生了一次重要的转变。由于他强调学习过程中的社会因素和认知过程在学习中的作用，因而在方法论上，班图拉必然注重以人为被试的实验。改变了行为主义以动物为实验对象，把由动物实验中得出的结论推广到人类学习现象的错误倾向。班图拉认为儿童通过观察他们生活中重要人物的行为而学得社会行为，这些观察以心理表象或其他符号表征的形式储存在大脑中，来帮助他们模仿行为。班图拉的这一理论接受了行为主义理论家们的大多数原理，但是更加注意线索对行为、对内在心理过程的作用，强调思想对行为和行为对思想的作用。他的观点在行为派和认知派之间架起一座桥梁，并对认知行为治疗作出了巨大的贡献。

班图拉的概念和理论建立在丰富坚实的实验验证资料的基础上，其实验方法比较严谨，结论比较有说服力。他的具有开放性的理论框架，在坚持行为主义立场的同时，积极吸取现代认知心理学的研究成果与研究方法，并受人本主义心理学若干思想的启发，涉及了观察学习、交互作用、自我调节、自我效能等重大课题，突出了人的主动性、社会性，受到心理学界的广泛赞同。认为个体、环境和行为是相互影响、彼此联系的。三者影响力的大小取决于当时的环境和行为的性质。在社会认知理论中，行为和环境都是可以改变的，但谁也不是行为改变的决定因素，例如攻击性强的儿童期望其他儿童对他产生敌意反应，这种期望使该儿童的攻击行为更有攻击性，从而又强化了该儿童的最初期望。

观察学习不要求必须有强化，也不一定产生外显行为。班图拉把观察学习分为以下四个过程。

### 1. 注意过程

注意和知觉榜样情景的各个方面。榜样和观察者的几个特征决定了观察学习的程度：观察者比较容易观察那些与他们自身相似的或者被认为是优秀的、热门的和有力的榜样。有依赖性的、自身概念低的或焦虑的观察者更容易产生模仿行为。强化的可能性或外在的期望影响个体决定观察谁、观察什么。

### 2. 保持过程

记住他们从榜样情景了解的行为，所观察的行为在记忆中以符号的形式表征，个体使用

两种表征系统——表象和言语。个体储存他们所看到的感觉表象,并且使用言语编码记住这些信息。

### 3. 复制过程

复制从榜样情景中所观察到的行为。个体将符号表征转换成适当的行为,个体必须:①选择和组织反应要素;②在信息反馈的基础上精炼自己的反应,即自我观察和矫正反馈。自我效能感是影响复制过程的一个重要因素,所谓自我效能感,即一个人相信自己能成功地执行产生一个特定的结果所要求的行为。如果学习者不相信自己能掌握一个任务,他们就不能继续做一个任务。

### 4. 动机过程

因表现所观察到的行为而受激励。观察学习论区分获得和表现,因为个体并不模仿他们所学的每一件事,强化非常重要,但并不是因为它增强行为,而是提供了信息和诱因,对强化的期望影响观察者注意榜样行为,激励观察者编码和记住可以模仿的、有价值的行为。

除了这种直接强化外,班图拉还提出了另外两种强化:替代性强化和自我强化。替代性强化指观察者因看到榜样受强化而受到的强化。例如当教师强化一个学生的助人行为时,班上的其他人也将花一定时间互帮互助。此外替代性强化还有一个功能,就是情绪反应的唤起。例如当电视广告上某明星因穿某种衣服或使用某种洗发精而风度迷人时,如果你感觉到或体验到因明星受到注意而感觉到的愉快,对于你这是一种替代性强化。自我强化依赖于社会传递的结果。社会向个体传递某一行为标准,当个体的行为表现符合甚至超过这一标准时,他就对自己的行为进行自我奖励。此外,班图拉还提出了自我调节的概念。班图拉假设,人们能观察他们自己的行为,并根据自己的标准进行判断,并由此强化或惩罚自己。

## 8.6 内省学习

内省是指对一个人自己的思想或情感进行考察,即自我观察;也指对自己在受控制的实验条件下进行的感觉和知觉经验所做的考察。内省是与外观相对的。外观是对自身以外的情况进行的研究和观察。内省法是早期心理学的一种研究方法,它根据被试者报告或描述的自己的体验来研究心理现象和过程。内省学习则是将内省概念引入机器学习中,即是通过检查和关心智能系统自身的知识处理和推理方式,从失败或低效中发现问题,形成修正自身的学习目标,由此改进自身处理问题方法的一种学习方式。

具备内省能力的学习系统也将提高学习效率。内省学习能使系统在分析执行任务成功和失败的基础上决定它的学习目标,而不是依靠系统设计者或用户给学习系统提供一个学习目标或目标概念。系统能明确地决定在什么地方出错的基础上需要学习什么。换而言之,内省学习系统能够理解在执行系统的运行中的失败及与之相关的系统推理和知识方面的原因。系统具有关于自己的知识和检查自己推理能力的本领,这样才能有效地学习。没有这种内省的愿望,学习是低效的。因此,对于有效学习内省是必要的。

内省学习可分为四个子问题:①有标准决定在什么时候检查推理过程,即监视推理过

程；②根据标准确定失败推理是否发生；③确定已检出失败的最终原因；④改变推理过程以免以后的类似失败。为了能发现和解释推理失败，内省学习系统需要能访问到关于系统本身推理过程直到当前时刻的知识。它需要粗略地或明确地知道关于领域内的结果和本身内部推理过程的期望。它需要能够在推理过程和问题解决执行中发现期望失败，还能够用根本推理失败解释期望失败并能决定为以后怎样改变推理过程来改正错误。内省学习过程可以包括三个步骤：判定一个失败，根据内省期望，确定是否生成期望失败；解释失败，内省学习系统检查失败之前的推理路线诊断出错原因，给出推理失败的结果解释，系统提出明确表示的内省学习目标用来改变它的知识和推理过程；修正失败，依据内省学习目标，系统构造并执行内省学习策略。可以看出，一个内省学习系统必须具有本身完整的学习结构，除知识库、算法库和推理机等以外，实现内省学习还需要一套表示系统推理过程的元推理表示法，用以实现跟踪和确定推理的实现过程；系统还需要一套推理评价标准，它包括期望、效率分析、出错、失败等一系列从现象到原因的分析和评估标准；同时还要有相应的改进推理的检错、目标形成和策略执行的内省机制。

Meta-AQUA 是库克斯(M T Cox)提出的一个应用解释模式(Explanation Pattern XP)用来解释异常的故事理解系统[162]。它是一个目标驱动学习系统，它从算法工具箱选择和结合多种学习方法修复导致系统失败的错误成分。系统输入是故事事件概念实体的数据流，执行任务是生成一个连贯解释和预见性的人物与事件交互模型。当系统预言某个解释将在故事中成立，而故事随后出现了一个相异的解释，这样一个失败就发生了。系统通过解释失败为什么发生，形成一个目标，并生成一个改变导致失败的知识学习策略，实现从失败中学习。

当 Meta-AQUA 没有成功理解一个故事片段时，它应用“内省元解释模式”(Introspective Meta-XPs, IMXPs)修正它的推理过程。一个 IMXP 描述了与之相关的推理失败和修正推理的可能学习策略。一个 IMXP 可看作一个可以与实际推理过程描述相匹配的模板，它确定是否发生推理失败，而实际推理过程是由“踪迹元解释模式”(Trace Meat-XPs, TMXPs)表示的。针对不同的推理失败，Meta-AQUA 归纳了失败症状种类、失败原因类型和学习目标类型，根据推理失败的症状确定推理失败的原因，根据失败原因确定相应的学习目标。

ROBBIE 是由福克斯(S Fox)提出的一个路径规划系统[230]。它的任务是在一组街道环境下和具备有限的地图知识以及少量初始样本路径的条件下，生成从一个地点至另一个地点的路径规划。它通过在一个模拟世界中执行这些规划来检验规划的质量。ROBBIE 是一个基于案例的系统。通过增加案例来增加它的世界知识：当案例库增加时，它将对地图有越来越好的了解。并行于基于案例的推理过程，ROBBIE 的内省单元监视规划器的推理，并将实际过程与基于案例推理过程理想性能的期望相比较。当期望失败发生时(或者是灾难性的或者是效率性的)，内省单元将暂停规划任务，并试图解释失败和修正系统。当具备必需的信息时，解释和修正任务将从暂停的地方重新开始。ROBBIE 系统的特点是将内省学习应用在基于案例推理的检索模块，实现案例索引的精化(Refining)。在出现检索失败时，能通过内省单元查找失败原因，修正检索过程。此外系统形成了将内省学习应用于整个基于案例的推理过程的框架。

### 8.6.1 内省学习一般模型

一般内省过程分为三个部分：判定失败、解释失败及修正失败[708]。判定失败：在对推理过程建立明确的和有限的期望的基础上，将期望与系统的实际推理执行过程相比较，由此发现差异。期望的和要求的行为与实际行为之间的差异即为期望失败。确定失败是否发生意味着系统具有一组明确的关于当前系统推理点的期望值。在系统推理过程中同时监视期望失败：当推理过程的每一步进行时，将有关结果与相关的期望对照，发现期望失败。解释失败：依据期望失败的标准和推理踪迹查找对失败的解释。在查明原因后，相应提出一个明确的对推理过程的改进建议以避免再次发生同样失败。修正失败：对于推理过程的修正措施可以附加在特定期望上，这样当一个期望失败产生时，附加的修正方法也可以同时提出。修正方法的描述并不能详细到怎样修改和修改什么，因此系统还应包括形成修正策略的机制。修正模块依据失败的描述和建议的修正方法，生成实际的修正策略并进行实际的修改。

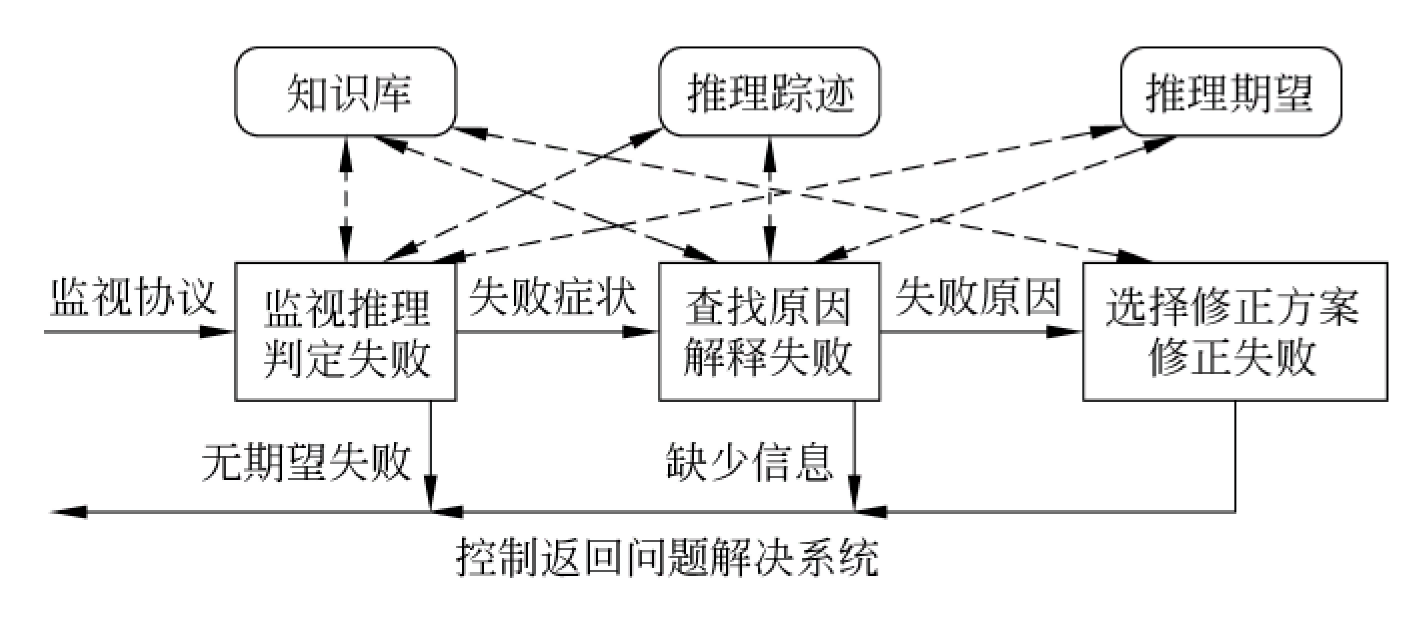



图 8.6 内省学习模型

图8.6为内省学习一般模型。模型除了包括判定失败、解释失败和修正失败三个过程以外，还包括知识库、推理踪迹、推理期望模型和监视协议等部分内容。监视协议规范怎样对系统推理过程进行监视。它规定在什么位置进行监视和如何监视以及系统控制权的转换。知识库包含系统推理相关知识，它不仅是系统推理的基础，同时也是判定和解释失败的依据。推理踪迹记录了系统推理过程，它专门用于内省学习，也是判定、解释和修正失败的重要依据。推理期望模型是系统推理过程的理想模型，它提供了推理期望的标准，因而是判定失败的主要依据。知识智能系统的内省学习单元依据监视协议，利用已有的背景知识、推理期望和推理踪迹检查当前状态是否发生期望失败。出现期望失败有两种情况：一种是当有一个关于推理过程的当前理想状态的模型期望与当前实际推理过程不符时，期望失败即产生；另一种是在系统发生灾难性失败而不能继续时，期望失败当然发生。如果推理单元没有发现期望失败，这意味着所有期望都和实际过程相符，系统将被通知一切正常，并重新获得控制权。如果发现了一个失败，推理单元将利用背景知识、推理踪迹和理想期望模型，查找失败的初始原因，解释失败。在一个失败被发现时，可得到的信息可能不足以诊断和修正这个失败。因此，内省学习单元可能暂停它的解释和修正任务允许系统继续工作直到有足够多的信息。当具备必需的信息时，解释和修正任务将从暂停的地方重新开始。失败的解释可为内省推理单元修正失败提供线索。失败解释后将生成修正失败的学习目标，修正失败模块将依据学习目标形成修正方案，并改变被确定是期望失败原因的推理。当修正完成后，或者发现修正不可能，系统将重新获得控制权。

### 8.6.2　内省学习的元推理

元推理源于认知过程中的元认知(Meta-cognition),也称自我监控认知。元认知是关于认知的认知。它指一个人关于自身认知过程和与之相关所有事物的知识。它包含觉知和控制两种能力。前者是指知道自己的能力限制、概念、知识与认知策略。后者是指适当的控制与使用自己的知识和策略的能力。以往的学习系统比较重视知识获取和处理,并且发展了知识保真系统,它强调知识的正确性。而在内省学习系统中,在保证知识的正确的同时,更重视推理过程的正确性。这就在知识表示的基础上提出对推理过程进行合理表示的要求。类似于数据仓库的元数据,即关于数据的数据,内省学习系统中利用了元推理,可以说是关于推理的推理。通过元推理的表示,学习系统能实现对知识和知识处理过程的认知,在此基础上实现适当地控制和使用自身的知识和策略。

元推理是关于推理的推理。因为内省学习的一个主要目标是依据推理失败或执行失败,修正推理过程,所以通过从根本层次表示推理是内省学习的一个基础条件。引入元推理需要达到两个目标,第一个目标是记录系统推理过程,形成推理踪迹,第二个目标是解释推理过程,提供推理失败的因果链。而最终目的是为监视推理过程提供表示方式,为解释和修正推理失败提供必要信息。

实现元推理的表示可用外部和内部两种方式。外部方式是对系统推理过程建立单独理想推理模型,在推理的不同阶段设计不同的评价标准,监督推理过程。内部方式是采用具备元解释功能的表示方式,从系统内部实现对推理过程的记录,并对异常进行解释。如 Meta-AQUQ 系统采用解释模式理论(XP)表示心智状态的转换,并命名这种表示结构为元解释模式。它的含意是一个关于另一个解释模式的解释模式。由于标准的解释模式本身就是一个因果结构的表示方法,而元解释模式更好地解释了本身解释过程的对错关系。依据功能的不同,系统提出了两种元解释模式,一种是用于跟踪解释过程的踪迹元解释模式(TMXP),另一种是用于内省解释过程的内省元解释模式(IMXP)。

### 8.6.3　失败分类

内省学习系统的另一个重要问题是失败分类。Meta-AQUA 系统分别列出了失败症状、失败原因和相应学习目标的类型。失败症状由实际值($A$)与期望值($E$)相同与否及存在与否分类,包括矛盾、意外、僵局、奇事、假命题、期望命题、退化等。每个症状由各种失败标记的组合表示。系统提出了两种错误标记即推论性期望失败和合并失败以及四种遗漏错误标记即误期预言、检索失败、构造失败和输入失败。ROBBIE 系统依据模型中不同模块将异常分类,并配以相应断言,而断言是解释失败的主要依据。失败由推理过程划分。推理过程分为索引案例、检索案例、调整案例、再检索案例、执行案例、保存案例等阶段,失败也相应分阶段划分。

失败分类是内省学习系统的一个要素。它是判定失败的基础,同时它也为解释失败和形成修正学习目标提供重要线索。失败分类在某种程度上还决定内省学习的能力。所以一个内省学习系统必需建立一个合理的失败分类。失败分类要考虑两个重要因素,一个方面

是失败分类的粒度,另一个方面是失败分类与失败解释及内省学习目标(修正失败)的关联性。对失败分层分类可以解决分类过细或过粗的矛盾。在失败大类中,可以抽象描述失败,还可以依据推理过程的不同阶段分大类,这样不仅可以包括一些不可预知的情况,增加系统内省的适应性,而且可以依据不同阶段,加快失败对照过程。细类可以较详细地描述失败,这样可以为失败解释提供有价值的线索。适当处理失败分类和失败解释的关联性也将提高系统内省能力。系统不仅需要依据失败症状方便推出失败原因,形成内省学习目标;系统还需要有处理各种不同问题的能力即适应性。失败解释同样可分为不同层次,抽象级和详细级或者多个层次。而失败分类的层次性也有助于形成合理的失败症状与失败解释的关系。

失败分类的方法可以用失败共性法和推理过程模块法。失败共性法是从失败的共同特征入手进行分类。例如将缺少输入信息归纳为输入失败,将推理机不能推理出或构造一个问题的解决方案归为构造失败,将知识的错误归为知识矛盾,推理性的错误归为推理失败等等。共性法是从系统整体方面考虑失败的分类。这种方法适合分布式环境下的内省学习。而模块法是将推理过程分为若干模块,按模块来划分失败。例如,将基于案例推理分为检索、调整、评估、保存等若干模块,检索失败指在检索过程中出现异常。模块法适合推理过程适用于模块化的系统,如模型选择等。在某些情况下,两种方法也可以结合运用。

### 8.6.4 内省过程中的基于案例推理

基于案例推理就是由目标案例的提示而获得记忆中的源案例,并由案案例来指导目标案例求解的一种策略。在基于范案推理中,把当前所面临的问题或情况称为目标案例,而把记忆中的问题或情况称为源案例。它的优势在于简化知识获取、提高求解效率和质量。基于案例推理的过程是依据当前目标形成案例检索特征,由检索特征从记忆案例库中检索,选择与当前情况最接近的案例,调整检索的案例适合当前情况,将当前事例形成新案例并评估新案例,最后将新案例保存入库以供以后使用。

基于案例推理和基于模型推理是实现内省学习的重要手段。同时内省学习也可以改进基于案例的推理过程。内省学习过程中的一个主要环节是依据失败特征,查找失败原因。基于案例推理适于这种匹配过程。内省不仅关注执行失败或者推理失败,还应涵盖低效的执行或者推理过程。内省学习系统除了发现错误以外,还需要对推理进行评价。从期望的角度看,判定失败也可称为监督与评价。将期望值归为监督与评价的标准,同时可以提出评估因子进行定量评估,监督是针对推理的过程,而评价针对推理结果。基于案例推理检索、调整、评价和保存的一系列过程实现判定失败和解释失败可以提高判定和解释效率,因此基于案例推理是一条有效途径。在 Meta-AQUA 系统中,从检错到形成学习目标这一过程就是一个基于案例推理的过程。系统通过失败症状查找失败原因并由此形成学习目标。而另一方面,将内省学习应用于基于案例推理的不同模块如检索和评价过程,则扩展了基于案例推理系统的适应能力和准确性。基于案例推理系统中案例评价是一个重要步骤,定量内省的案例评价可以使案例能依据用户偏好自动改变案例权值,提高案例检索效率。

# 8.7　强化学习

## 8.7.1　强化学习模型

强化学习不是通过特殊的学习方法来定义的，而是通过在环境中和响应外界环境的动作来定义的。任何解决这种交互的学习方法都是一个可接受的强化学习方法。强化学习也不是监督学习，在有关机器学习的部分我们都可以看出来。在监督学习中，“教师”用实例来直接指导或者训练学习程序。在强化学习中，学习智能体自身通过训练、误差和反馈，学习在环境中完成目标的最佳策略。

强化学习技术是从控制理论、统计学、心理学等相关学科发展而来的，最早可以追溯到巴甫洛夫的条件反射实验。但直到 20 世纪 80 年代末、90 年代初强化学习技术才在人工智能、机器学习和自动控制等领域中得到广泛研究和应用，并被认为是设计智能系统的核心技术之一。

强化学习的模型如图 8.7 所示，通过智能体与环境的交互进行学习。智能体与环境的交互接口包括行动(Action)、奖励(Reward)和状态(State)。交互过程可以表述为如下形式：每一步，智能体根据策略选择一个行动执行，然后感知下一步的状态和即时奖励，通过经验再修改自己的策略。智能体的目标就是最大化长期奖励。

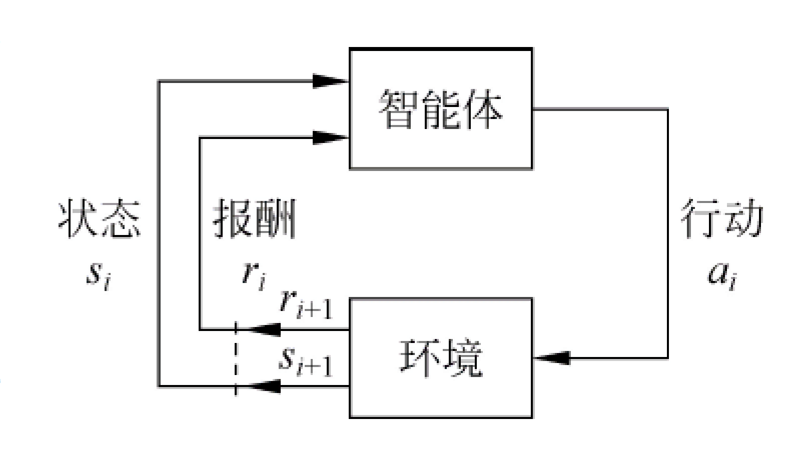


图 8.7　强化学习模型

强化学习系统接受环境状态的输入 $s$，根据内部的推理机制，系统输出相应的行为动作 $a$。环境在系统动作作用 $a$ 下，变迁到新的状态 $s'$。系统接受环境新状态的输入，同时得到环境对于系统的瞬时奖惩反馈 $r$。对于强化学习系统来讲，其目标是学习一个行为策略 $\pi$：$S \to A$，使系统选择的动作能够获得环境奖励的累计值最大。在学习过程中，强化学习技术的基本原理是：如果系统某个动作导致环境正的奖励，那么系统以后产生这个动作的趋势便会加强。反之系统产生这个动作的趋势便减弱。这和生理学中的条件反射原理是接近的。

$$\sum_{i=0}^{\infty} \gamma^i r_{t+i} \quad 0 < \gamma \leqslant 1 \tag{8.6}$$

如果假定环境是马尔科夫型的，则顺序型强化学习问题可以通过马尔可夫决策过程建模。下面首先给出马尔可夫决策过程的形式化定义。

马尔科夫决策过程由四元组 $<S,A,R,P>$ 定义。包含一个环境状态集 $S$，系统行为集合 $A$，奖励函数 $R$：$S \times A \to \mathcal{R}$ 和状态转移函数 $P$：$S \times A \to PD(S)$。记 $R(s,a,s')$ 为系统在状态 $s$ 采用 $a$ 动作使环境状态转移到 $s'$ 获得的瞬时奖励值；记 $P(s,a,s')$ 为系统在状态 $s$ 采用 $a$ 动作使环境状态转移到 $s'$ 的概率。

马尔科夫决策过程的本质是：当前状态向下一状态转移的概率和奖励值只取决于当前状态和选择的动作，而与历史状态和历史动作无关。因此在已知状态转移概率函数 $P$ 和奖励函数 $R$ 的环境模型知识下，可以采用动态规划技术求解最优策略。而强化学习着重研究在 $P$ 函数和 $R$ 函数未知的情况下，系统如何学习最优行为策略。

为解决这个问题，图 8.8 中给出强化学习四个关键要素之间的关系，即策略 $\pi$、状态值映射 $V$、奖励函数 $R$ 和一个环境的模型(通常情况)。四要素关系自底向上呈金字塔结构。策略定义在任何给定时刻学习智能体的选择和动作的方法。这样，策略可以通过一组产生式规则或者一个简单的查找表来表示。像刚才指出的，特定情况下的策略可能也是广泛搜索，查询一个模型或计划过程的结果。它也可以是随机的。策略是学习智能体中重要的组成部分，因为它自身在任何时刻足以产生动作。

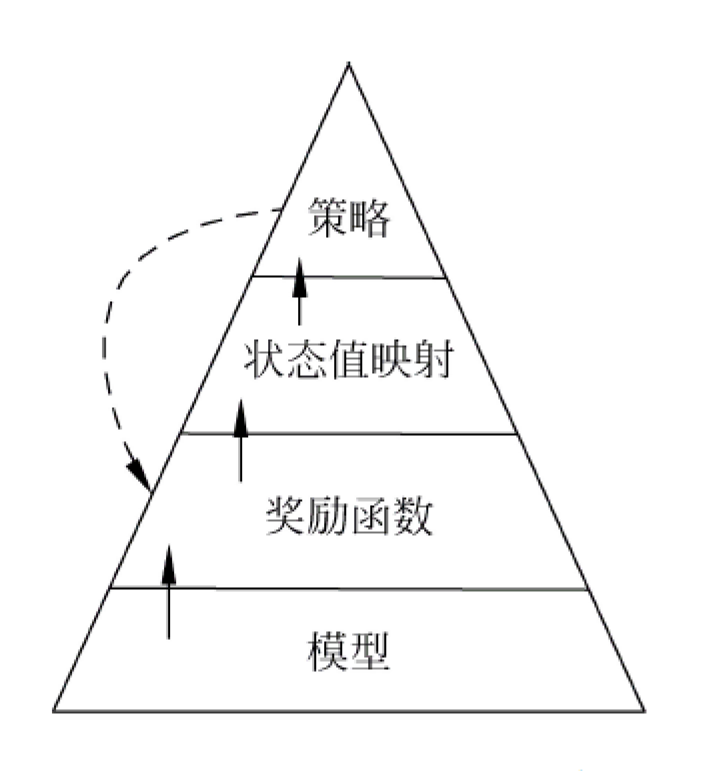



图 8.8 强化学习四要素

奖励函数 $R_t$ 定义了在时刻 $t$ 问题的状态/目标关系。它把每个动作，或更精细的每个状态-响应对，映射为一个奖励量，以指出那个状态完成目标的愿望的大小。强化学习中的智能体有最大化总的奖励的任务，这个奖励是它在完成任务时所得到的。

状态值映射 $V$ 是环境中每个状态的一个属性，它指出对从这个状态继续下去的动作系统可以期望的奖励。奖励函数度量状态-响应对的立即的期望值，而赋值函数指出环境中一个状态的长期的期望值。一个状态从它自己内在的品质和可能紧接着它的状态的品质来得到值，也就是在这些状态下的奖励。例如，一个状态/动作可能有一个低的立即的奖励，但有一个较高的值，因为通常紧跟它的状态产生一个较高的奖励。一个低的值可能同样意味着状态不与成功的解路径相联系。

如果没有奖励函数，就没有值，估计值的唯一目的是为了获取更多的奖励。但是，在做决定时，是值最使我们感兴趣，因为值指出带来最高的回报的状态和状态的综合。但是，确定值比确定奖励困难。奖励由环境直接给定，而值是估计得到的，然后随着时间推移根据成功和失败重新估计值。事实上，强化学习中最重要也是最难的方面是创建一个有效的确定值的方法。

强化学习的环境模型是抓住环境行为的方面的一个机制。模型让我们在没有实际试验它们的情况下估计未来可能的动作。基于模型的计划是强化学习案例的一个新的补充，因为早期的系统趋向于基于纯粹的一个智能体的试验和误差来产生奖励和值参数。

系统所面临的环境由环境模型定义，但由于模型中 $P$ 函数和 $R$ 函数未知，系统只能够依赖于每次试错所获得的瞬时奖励来选择策略。但由于在选择行为策略过程中，要考虑到环境模型的不确定性和目标的长远性，因此在策略和瞬时奖励之间构造值函数(即状态的效用函数)，用于策略的选择。

首先通过式(8.7)构造一个返回函数 $R_t$，用于反映系统在某个策略 $\pi$ 指导下的一次学习循环中，从 $s_t$ 状态往后所获得的所有奖励的累计折扣和。由于环境是不确定的，系统在某个策略 $\pi$ 指导下的每一次学习循环中所得到的 $R_t$ 有可能是不同的。因此在 $s$ 状态下的值函数要考虑不同学习循环中所有返回函数的数学期望。因此在 $\pi$ 策略下，系统在 $s$ 状态下的值函数由式(8.8)定义，其反映了如果系统遵循 $\pi$ 策略，所能获得的期望的累计奖励折扣和。

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = r_{t+1} + \gamma R_{t+1} \tag{8.7}$$

$$V^\pi(s) = E_\pi\{R_t \mid s_t = s\} = E_\pi\{r_{t+1} + \gamma V(s_{t+1}) \mid s_t = s\} = \sum_a \pi(s,a) \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma V^\pi(s')] \tag{8.8}$$

根据 Bellman 最优策略公式，在最优策略 $\pi^*$ 下，系统在 $s$ 状态下的值函数由(8.9)式定义。

$$V^*(s)=\max_{a\in A(s)}E\{r_{t+1}+\gamma V^*(s_{t+1})\mid s_t=s,a_t=a\}=\max_{a\in A(s)}\sum_{s'}P_{ss'}^{a}[R_{ss'}^{a}+\gamma V^*(s')] \tag{8.9}$$

在动态规划技术中，在已知状态转移概率函数 $P$ 和奖励函数 $R$ 的环境模型知识前提下，从任意设定的策略 $\pi_0$ 出发，可以采用策略迭代的方法(式(8.10)和式(8.11))逼近最优的 $V^*$ 和 $\pi^*$。式(8.10)和式(8.11)中的 $k$ 为迭代步数。

$$\pi_k(s)=\arg\max_{a}\sum_{s'}P_{ss'}^{a}[R_{ss'}^{a}+\gamma V^{\pi_{k-1}}(s')] \tag{8.10}$$

$$V^{\pi_k}(s)\leftarrow\sum_{a}\pi_{k-1}(s,a)\sum_{s'}P_{ss'}^{a}[R_{ss'}^{a}+\gamma V^{\pi_{k-1}}(s')] \tag{8.11}$$

但由于强化学习中，$P$ 函数和 $R$ 函数未知，系统无法直接通过式(8.10)、式(8.11)进行值函数计算。因而实际中常采用逼近的方法进行值函数的估计，其中最主要的方法之一是蒙特卡罗(Monte Carlo)采样，如式(8.12)。其中 $R_t$ 是指当系统采用某种策略 $\pi$，从 $s_t$ 状态出发获得的真实的累计折扣奖励值。保持 $\pi$ 策略不变，在每次学习循环中重复地使用式(8.12)，下式将逼近(8.8)式。

$$V(s_t)\leftarrow V(s_t)+\alpha[R_t-V(s_t)] \tag{8.12}$$

结合蒙特卡罗方法和动态规划技术，式(8.13)给出强化学习中时间差分学习(Temporal Difference，TD)的值函数迭代公式。

$$V(s_t)\leftarrow V(s_t)+\alpha[r_{t+1}+\gamma V(s_{t+1})-V(s_t)] \tag{8.13}$$

## 8.7.2　Q 学习

在 Q 学习中，Q 是状态-动作对到学习到的值的一个函数。对所有的状态和动作：

Q: (state x action) → value

对 Q 学习中的一步：

$$Q(s_t,a_t)\leftarrow(1-c)\times Q(s_t,a_t)+c\times[r_{t+1}+\gamma\,\mathrm{MAX}_{a}Q(s_{t+1},a)-Q(s_t,a_t)] \tag{8.14}$$

其中 $c$ 和 $\gamma$ 都 $\leqslant 1$，$r_{t+1}$ 是状态 $s_{t+1}$ 的奖励。

在 Q 学习中，回溯从动作结点开始，最大化下一个状态的所有可能动作和它们的奖励。在完全递归定义的 Q 学习中，回溯树的底部结点一个从根结点开始的动作和它们的后继动作的奖励的序列可以到达的所有终端结点。联机的 Q 学习，从可能的动作向前扩展，不需要建立一个完全的世界模型。Q 学习还可以脱机执行。我们可以看到，Q 学习是一种时序差分的方法。

**算法 8.1**　Q 学习算法

```
Initialize Q(s,a) arbitrarily
Repeat (for each episode)
  Initialize s
  Repeat (for each step of episode)
  Choose a from s using policy derived from Q (e.g.,ε-greedy)
   Take action a, observer r, s'
     Q(s,a)←Q(s,a) + α[r + γmax_a' Q(s',a') - Q(s,a)]
```

```
        s←s'
    Until s is terminal
```

### 8.7.3 部分感知强化学习

在实际应用中,学习系统往往难以完全准确地观察到环境的真实状态,而只能观察到真实状态的某一个或某几个方面。这种状态观察上的不确定性为动作评估带来了更多的不确定性,从而直接影响所选动作的好坏。图8.9给出了部分感知状态马尔科夫模型。在经典的马尔科夫模型上增加状态预测,并对每个状态设置一个信度b,用于表示该状态的可信度,在决定动作时使用b作为依据,同时根据观察值进行状态预测,这样能解决一些非马尔科夫模型的问题。

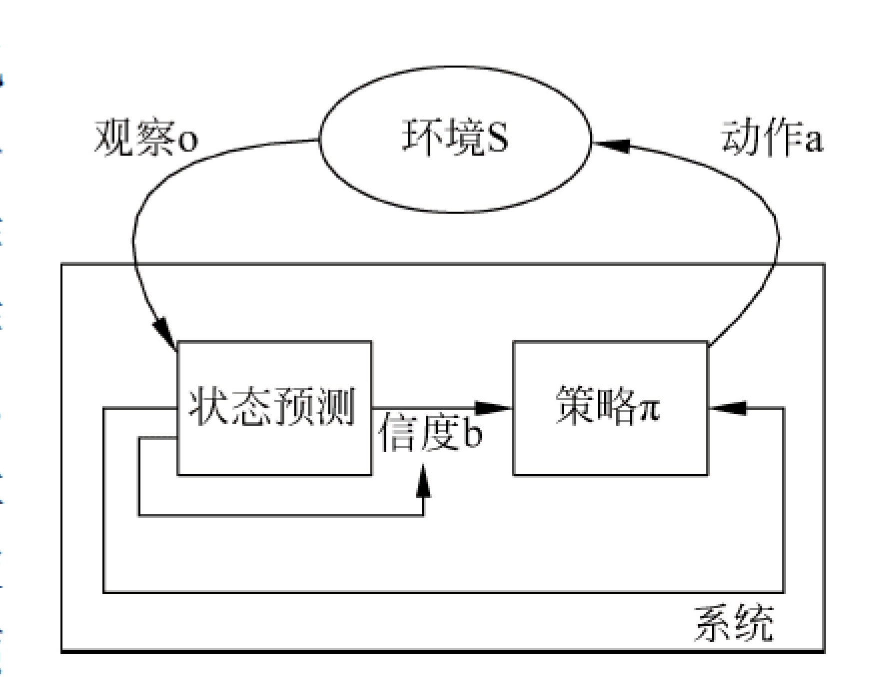



图8.9 部分感知状态马尔科夫模型

由于部分感知问题的核心是观测状态的不确定性,因此部分感知强化学习研究的核心是在学习过程中消除不确定性。理论上,这种不确定性可以通过概率来表示(信度),然后构建基于信度的马氏决策过程。但由于实际中信度是一个连续值,问题转成大规模顺序决策任务,需要用函数估计强化学习技术来解决,效果并不理想[356]。实际解决方法采用以下思路。

(1) 通过一系列历史的观测来构造状态。使所构造的状态满足马尔科夫属性。该类方法的代表有K-历史窗口方法[435]。

(2) 通过对观测分析,如通过预测下一观测的能力、预测行为的奖赏,对观测进行分割,以期分割成实际的状态。该类方法的代表有NSM和USM方法[473]。

在预测状态表示模型中,无须系统隐藏的实际状态,而利用动作-观测值序列来构建系统的模型[447]。基于预测状态表示模型的规划和学习比基于部分感知马氏决策模型更有优势。

## 8.8 深度学习

### 8.8.1 概述

1981年的诺贝尔医学奖,颁发给了休伯(David Hubel)和威塞尔(Torsten Wiesel),以及斯佩里(Roger Sperry)。前两位的主要贡献,是发现了视觉系统的信息处理:可视皮质是分级的,如图8.10所示。从低级的V1区提取边缘特征,再到V2区的形状或者目标的部分等,再到更高层,整个目标、目标的行为等。

深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,以发现数据的分布式特征表示。含多隐层的多层感知器就是一种深度学习结构。深度学习是机器学习研究中的一个新的领域,其核心思想在于模拟人脑的层级抽象结构,通过无监督的方式分析大规模数据,发掘大数据中蕴藏的有价值信息。深度学习应大数据而生,给大数据提供了一个深度思考的大脑。[942]

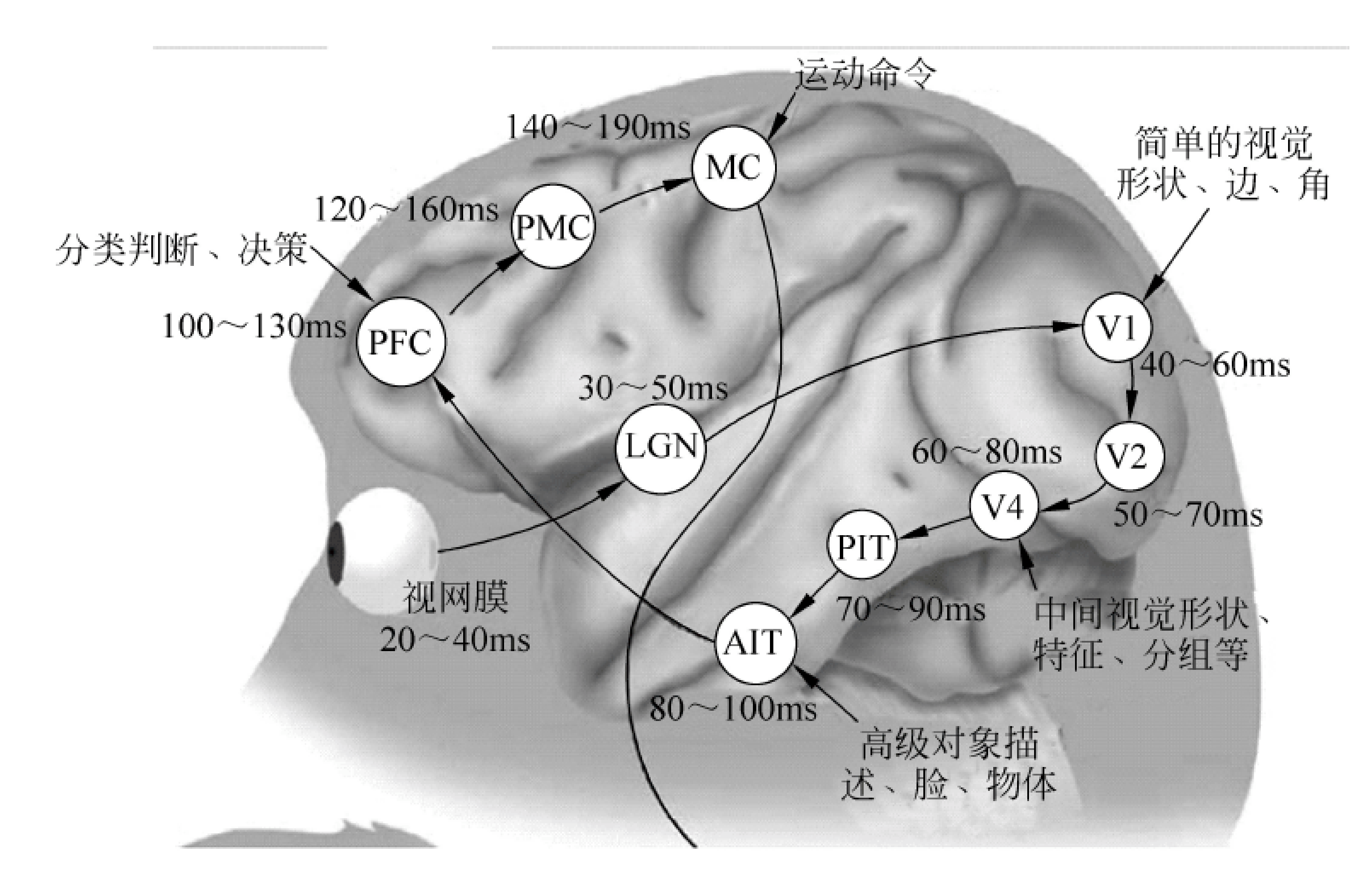



图 8.10 人脑视觉系统

深度学习的概念由 Hinton 等人于 2006 年提出[320]。基于深度置信网络(DBN)提出非监督贪心逐层训练算法[319]，为解决深层结构相关的优化难题带来希望，随后提出多层自动编码器深层结构。杨立昆(Y Lecun)等人提出的卷积神经网络是第一个真正多层结构学习算法[408]，它利用空间相对关系减少参数数目以提高训练性能。

## 8.8.2 深度信念网络

受限玻尔兹曼机是一个单层的随机神经网络(通常我们不把输入层计算在神经网络的层数里)，本质上是一个概率图模型。输入层与隐层之间是全连接，但层内神经元之间没有相互连接。每个神经元要么激活(值为 1)要么不激活(值为 0)，激活的概率满足 sigmoid 函数。RBM 的优点是给定一层时另外一层是相互独立的，那么做随机采样就比较方便，可以分别固定一层，采样另一层，交替进行。

2006 年，辛顿等在文献[319]中，提出了一种深度信念网络(Deep Belief Nets，DBN)，如图 8.11 所示。一个深度神经网络模型可被视为由若干个 RBM 堆叠在一起，这样一来，在训练的时候，就可以通过由低到高逐层训练这些 RBM 来实现。由于 RBM 可以通过 CD 算法进行快速训练，因此，这一框架绕过了直接从整体上训练深度神经网络的高度复杂性，而将其化简为对多个 RBM 的训练问题。辛顿建议，经过这种方式训练后，可以再通过传统的全局学习算法(例如反向传播、Wake-Sleep 算法)对网络进行微调，从而使模型收敛到一个局部最优点上。这种学习算法，本质上等同于先通过逐层 RBM 训练将模型的参数初始化为一个较优的值，然后再通过少量的传统学习算法进一步训练。这样一方面解决了模型训练慢的问题，另一方面，大量实验也证明，这种方式能够产生非常好的参数初始化值，从而也提升了最终参数的质量。

图 8.12 的左边图给出一个例子[319]。这个网络有 4 层，将一个高维的图像信号压缩到 30 维，即最顶层的神经元个数为 30。我们还可以将这个网络对称展开，从 30 维回到原来的高维信号，这样就有了一个 8 层的网络(见图 8.12 的中间图)。如果该网络用于信号压缩，那么可以令该网络的目标输出等于输入，再用 BP 算法对权值进行微调(见图 8.12 的右边图)。

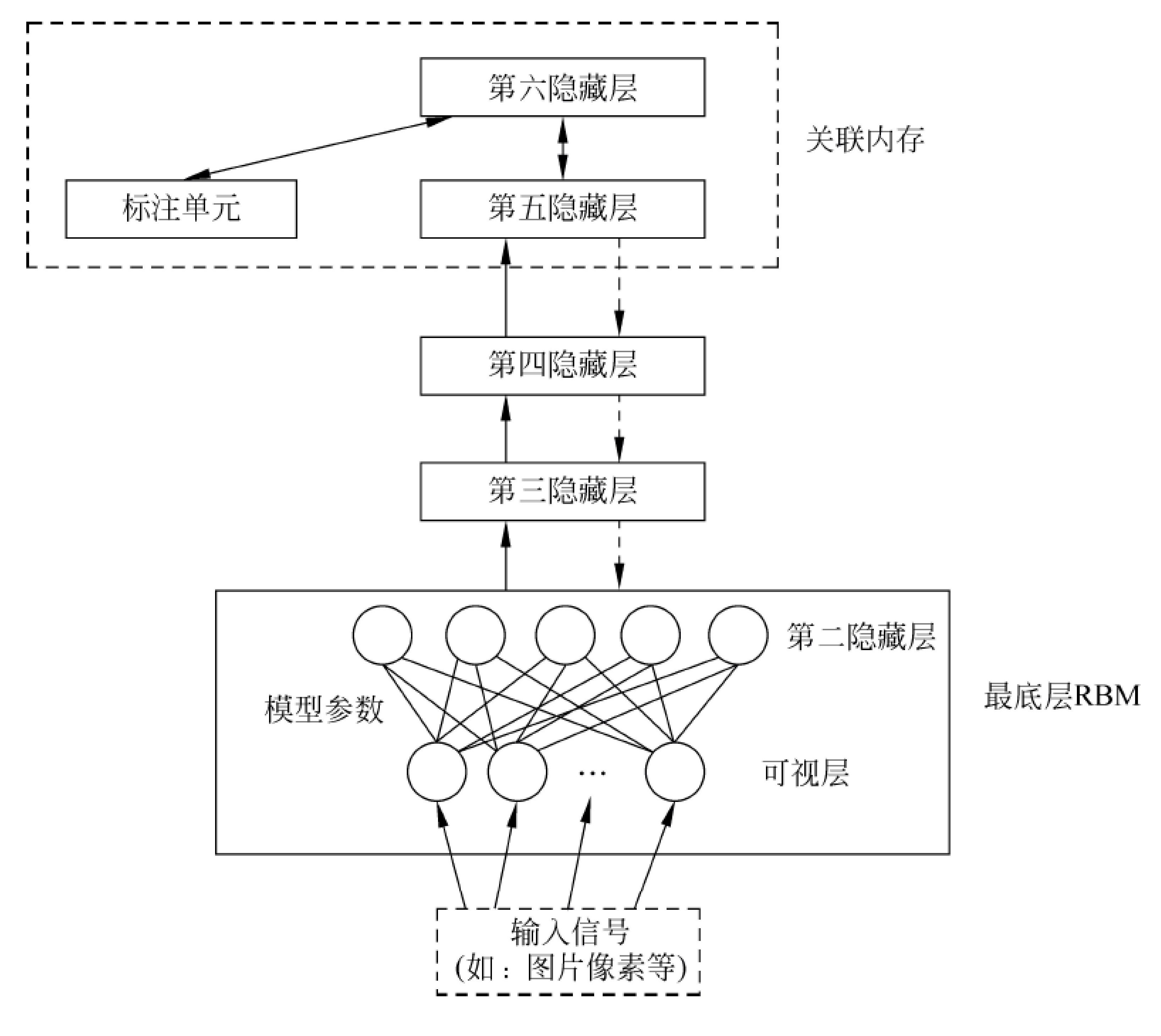



图 8.11　深度信念网络

## 8.8.3　卷积神经网络

卷积神经网络是一种多阶段、全局可训练的人工神经网络模型[402]，它可以从经过少量预处理、甚至原始数据中学习到抽象的、本质的和高阶的特征，在车牌检测、人脸检测，手写体识别、目标跟踪等领域得到了广泛的应用。

卷积神经网络在二维模式识别问题上，通常表现的比多层感知器好，原因在于卷积神经网络在结构中加入了二维模式的拓扑结构，并使用三种重要的结构特征：局部接受域、权值共享和子采样来保证输入信号的目标平移、放缩和扭曲一定程度上的不变性。卷积神经网络主要由特征提取和分类器组成，特征提取包含多个卷积层和子采样层，分类器一般使用一层或两层的全连接神经网络。卷积层具有局部接受域结构特征，子采样层具有子采样结构特征，这两层都具有权值共享结构特征。图 8.13 是一个用于手写体识别的卷积神经网络的结构[409]。

图 8.13 中，卷积神经网络共有七层：一个输入层，两个卷积层，两个子采样层和两个全连接层。输入层每个输入样本包含 32×32=1024 个像素。C1 为卷积层，包含 6 个特征图，每个特征图包含 28×28=784 个神经元。C1 上每个神经元通过 5×5 的卷积核与输入层相应 5×5 的局部接受域相连，卷积步长为 1，所以 C1 层共包含 6×784×(5×5+1)=122 304 个连接。每个特征图包含 5×5 个权值和一个偏置，所以 C1 层共包含 6×(5×5+1)=156 个可训练参数。

S1 为子采样层，包含 6 个特征图，每个特征图包含 14×14=196 个神经元。S1 上的特征图与 C1 层上的特征图一一对应，子采样窗口为 2×2 的矩阵，子采样步长为 1，所以 S2 层

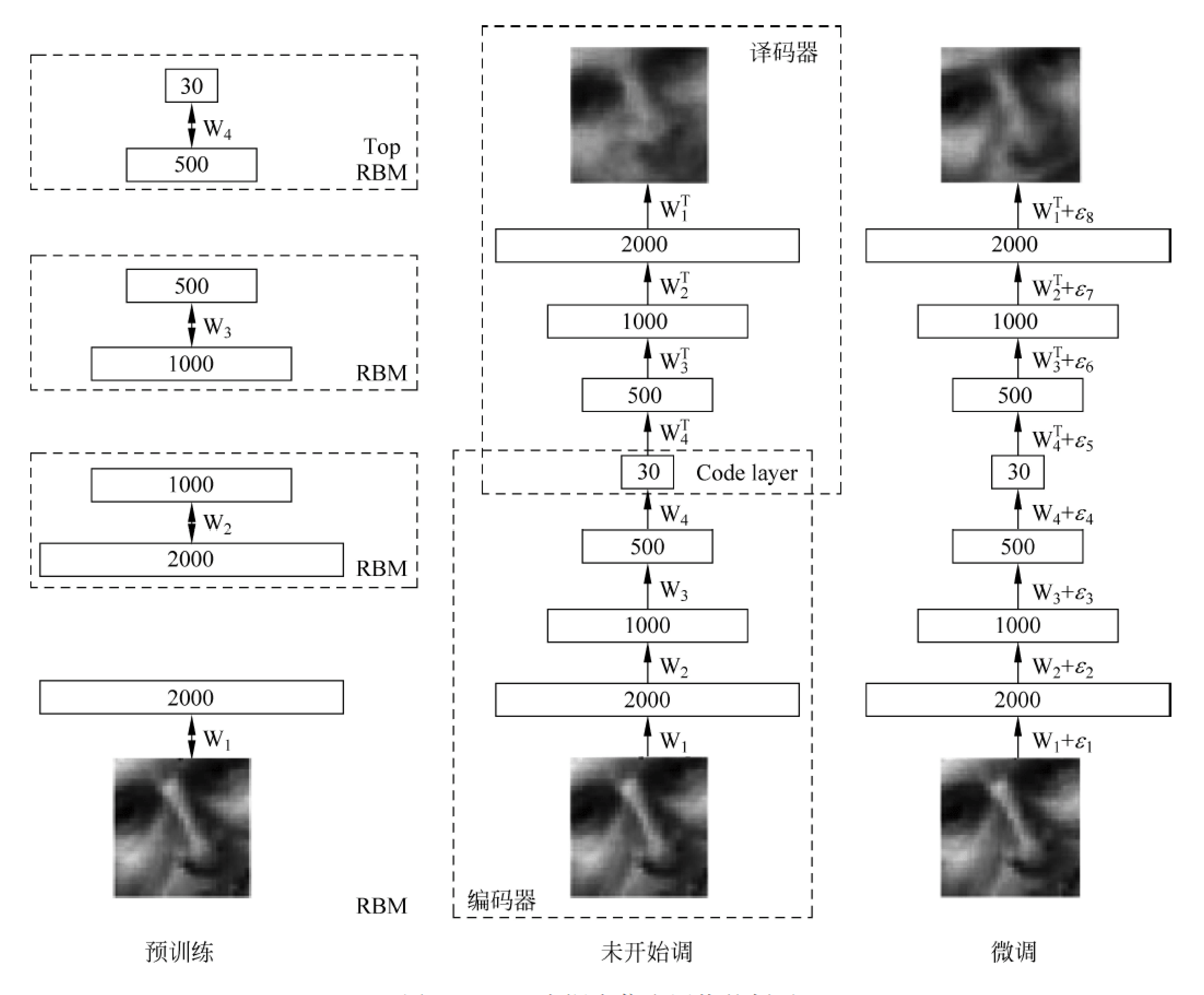



图 8.12　一个深度信念网络的例子

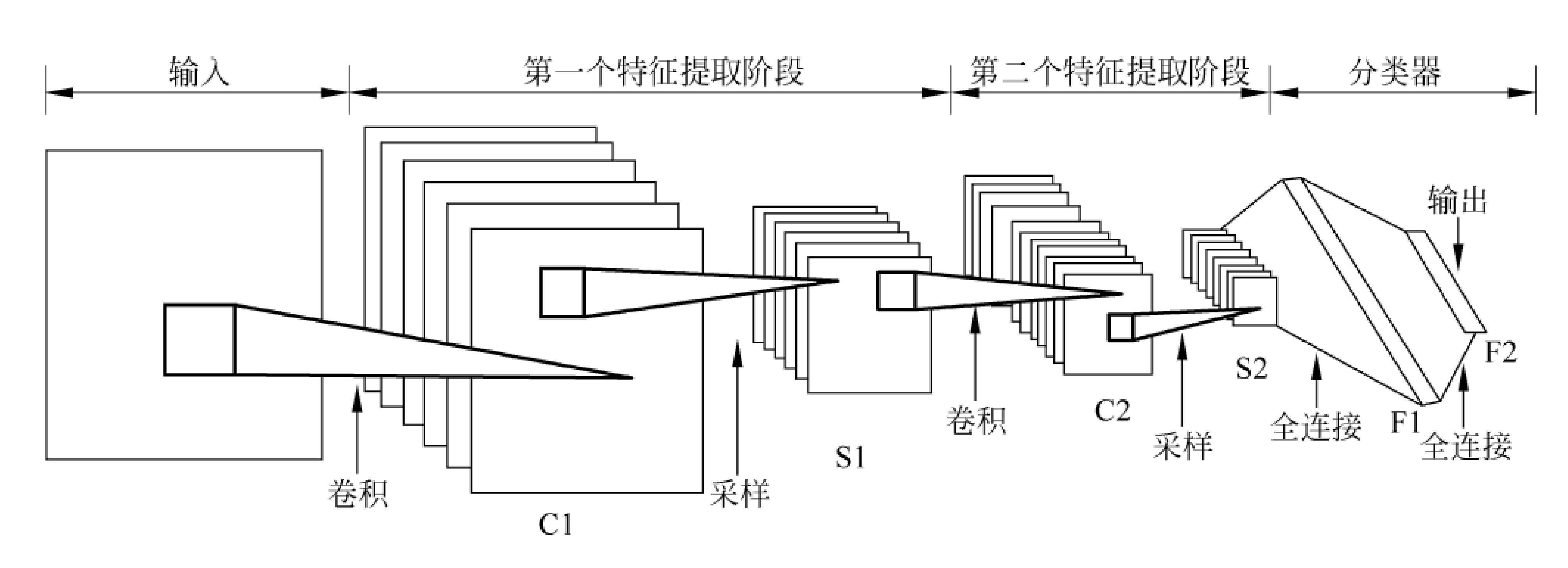



图 8.13　手写体识别的卷积神经网络的结构示意图

共包含 6×196×(2×2+1)=5 880 个连接。S1 上的每个特征图含有一个权值和一个偏置，所以 S2 层共有 12 个可训练参数。

C2 为卷积层，包含 16 个特征图，每个特征图包含 10×10=100 个神经元。C2 上每个神经元通过 $k$ 个($k\leqslant 6$,6 为 S1 层上的特征图个数)5×5 的卷积核与 S1 上 $k$ 个特征图中相应 5×5 的局部接受域相连。使用全连接的方式时 $k=6$。所以实现的卷积神经网络 C2 层共包含 41 600 个连接。每个特征图包含 6×5×5=150 个权值和一个偏置，所以 C1 层共包

含 16×(150+1)=2416 个可训练参数。

S2 为子采样层,包含 16 个特征图,每个特征图包含 5×5 个神经元,S2 共包含 400 个神经元。S1 上的特征图与 C2 层上的特征图一一对应,S2 上特征图的子采样窗口为 2×2,所以 S2 层共包含 16×25×(2×2+1)=2000 个连接。S2 上的每个特征图含有一个权值和一个偏置,所以 S2 层共有 32 个可训练参数。

F1 为全连接层,包含 120 个神经元,每个神经元都与 S2 上 400 个神经元相连,所以 F1 包含连接数与可训练参数都为 120×(400+1)=48 120。F2 为全连接层,也是输出层,包含 10 个神经元、1210 个连接和 1210 个可训练参数。

从图 8.13 中可以看出,卷积层特征图数目逐层增加,一方面是为了补偿采样带来的特征损失,另一方面,由于卷积层特征图是由不同的卷积核与前层特征图卷积得到,即获取的是不同的特征,这就增加了特征空间,使提取的特征更加全面。

卷积神经网络在有监督的训练中多使用误差反向传播(BP)算法,采用基于梯度下降的方法,通过误差反向传播不断调整网络的权值和偏置,使训练集样本整体误差平方和最小。BP 训练算法可以分为四个过程:网络初始化,信息流的前向传播,误差反向传播,权值和偏置更新。在误差反向逐层传递过程中,还需计算权值和偏置的局部梯度改变量。

在训练阶段的开始,需要为各层神经元随机初始化权值。权值的初始化对网络的收敛速度有很大影响,所以如何初始化权值是非常重要的。权值的初始化与网络选取的激活函数有关,为了加快收敛速度,权值尽量取到激活函数变化最快的部分,初始化的权值太大或太小都导致权值的变化量很小。

在信息流的前向传播中,卷积层首先提取输入中的初级基本特征,形成若干特征图,然后子采样层降低特征图的分辨率。卷积层和子采样层交替完成特征提取阶段之后,这时,网络获取了输入中的高阶的不变性的特征。然后,这些高阶的不变性特征前向反馈到全连接神经网络,由全连接神经网络对这些特征进行分类。经过全连接神经网络隐藏层和输出层信息变换和计算处理,就完成了一次学习的正向传播处理过程,最终结果由输出层向外界输出。

当实际输出与期望输出不符合时,网络进入误差反向传播阶段。误差从输出层传递到隐含层,从隐含层再传递到特征提取阶段的子采样层和卷积层。各层神经元都获取到自己的输出误差之后,开始计算每个权值和偏置的局部改变量,最后进入到权值更新阶段。

### 1. 卷积层前向传播

卷积层的每一个神经元提取前一层全部特征图中相同位置局部接受域中的特征,并且,同一特征图上的神经元共享一个权值矩阵。卷积过程可看做卷积层神经元通过权值矩阵对前层特征图逐行逐列无缝扫描。第 $l$ 层卷积层第 $k$ 个特征图中第 $x$ 行第 $y$ 列的神经元的输出 $\mathrm{O}_{(x,y)}^{(l,k)}$ 可由式(8.15)求得,其中,$\tanh(\cdot)$是激活函数。

$$\mathrm{O}_{(x,y)}^{(l,k)} = \tanh\left(\sum_{t=0}^{f-1}\sum_{r=0}^{\mathrm{kh}}\sum_{c=0}^{\mathrm{kw}} \mathrm{W}_{(r,c)}^{(k,t)}\mathrm{O}_{(x+r,y+c)}^{(1-l,t)} + \mathrm{Bias}^{(l,k)}\right) \tag{8.15}$$

从式(8.15)可以看出,计算卷积层一个神经元的输出需要遍历前层各个特征图中相应卷积窗口中的神经元。全连接层的前向传播与卷积层前向传播类似,可以看作卷积权值矩阵与输入大小相同的卷积操作。

### 2. 子采样层前向传播

子采样层的特征图与前层卷积层特征图数目相同且一一对应。每一个神经元通过子采样窗口与前层相应特征图中大小相同但互不重叠的子区域相连。第 $l$ 层子采样层第 $k$ 个特征图中第 $x$ 行第 $y$ 列神经元的输出 $O_{(x,y)}^{(l,k)}$ 为式(8.16)所示。

$$O_{(x,y)}^{(l,k)} = \tanh\left(W^{(k)}\sum_{r=0}^{sh}\sum_{c=0}^{sw}O_{(xsh+r,ysw+c)}^{(l-1,t)} + \mathrm{Bias}^{(l,k)}\right) \tag{8.16}$$

### 3. 子采样层误差反向传播

误差反向传播从输出层开始，经过隐藏层传入子采样层。输出层的误差反向传播首先计算误差关于输出层神经元输出的偏导数。假设训练样本 $d$ 在输出层第 $k$ 个神经元的输出为 $o_k$，而样本 $d$ 的在输出层第 $k$ 个输出单元的期望输出为 $t_k$，那么样本 $d$ 在输出层的误差 $E=1/2\Sigma_k(o_k-t_k)^2$。误差 $E$ 关于输出 $o_k$ 的偏导数为$\partial E/\partial o_k=o_k-t_k$，类似可以求得误差关于输出层所有神经元的偏导数。之后，需要求出误差关于输出层各神经元输入的偏导数。设输出层误差关于第 $k$ 个神经元输入的偏导数为 $d(o_k)$，可由式(8.17)得出，其中$(1+o_k)(1-o_k)$为激活函数 $\tanh(\cdot)$对该神经元的输入求偏导，然后开始计算误差关于隐层各神经元输出的偏导数。假设隐层的一个神经元为 $j$，$w_{kj}$ 为神经元 $j$ 与输出层神经元之间的连接权值，那么误差关于神经元 $j$ 的输出的偏导数 $d(o_j)$可由式(8.18)求得。将误差关于隐层各神经元的输出的偏导数求出之后，误差就从输出层反向传播到了隐层，隐层经过相似的过程将误差反向传播到子采样层。为表述方便，这里将误差关于神经元输出的偏导数称为神经元的输出误差，将误差关于神经元输入的偏导数称为神经元的输入误差。

$$d(o_k) = (o_k - t_k)(1 + o_k)(1 - o_k) \tag{8.17}$$

$$d(o_j) = \Sigma \mathrm{d}(o_k)w_{kj} \tag{8.18}$$

子采样层特征图个数与卷积层特征图个数相同且一一对应，所以误差从子采样层传播到卷积层比较直观。先利用式(8.17)计算子采样层各神经元的输入误差，然后将神经元输入误差传播到子采样层的前层神经元。设子采样层为第 $l$ 层，则第 $l-1$ 层第 $k$ 个特征图中第 $x$ 行第 $y$ 列神经元的输出误差可由式(8.19)所得。

$$d\left(O_{(x,y)}^{(l-l,k)}\right) = d\left(O_{(\lfloor x/xh\rfloor,\lfloor y/\ xw\rfloor)}^{(l,k)}\right)W^{(k)} \tag{8.19}$$

子采样层的一个特征图中所有神经元共享一个权值和一个偏置，所以权值和偏置的局部梯度改变量与子采样层所有神经元相关。子采样层 $l$ 第 $k$ 个特征图的权值改变量 $\Delta W^{(k)}$ 和偏置改变量 $\Delta\mathrm{Bias}^{(l,k)}$ 分别可由式(8.20)和式(8.21)所得。在式(8.21)中，$f_h$ 和 $f_w$ 代表子采样层 $l$ 中特征图的高和宽。

$$\Delta W^{(k)} = \sum_{x=0}^{fn}\sum_{y=0}^{fw}\sum_{t=0}^{sh}\sum_{c=0}^{sw}O_{(x,y)}^{(l-l,k)}d\left(O_{(\lfloor x/sh\rfloor,\lfloor y,sw\rfloor)}^{(l,t)}\right) \tag{8.20}$$

$$\Delta\mathrm{Bias}^{(l,k)} = \sum_{x=0}^{fh}\sum_{y=0}^{fw}d\left(O_{(x,y)}^{(l,k)}\right) \tag{8.21}$$

#### 4. 卷积层误差反向传播

卷积层的误差反向传播主要有“推”和“拉”两种方式。“推”的方式(见图 8.14(a))可理解为卷积层神经元主动将误差传给前层神经元,这种方式比较适合串行实现,但在并行实现时存在“写冲突”问题;“拉”的方式(见图 8.14(b))可理解为后层神经元主动从前层各神经元获取误差,这种方式实现起来比较复杂,由于卷积操作的边缘效应,需要先确定当前层神经元与前层特征图中的哪些神经元相连。

采用“拉”的方式描述卷积层误差反向传播过程。先利用式(8.17)计算卷积层各神经元的输入误差,然后将神经元输入误差传播到卷积层的前层神经元,如式(8.22)。

$$d\left(\mathrm{O}_{(x,y)}^{(l-1,k)}\right)=\sum_{t=0}^{m-1}\sum_{(p,q)\in A}d\left(\mathrm{O}_{p,q}^{(l,t)}\right)w \tag{8.22}$$

其中,$A$ 为第 $l-1$ 层第 $k$ 个特征图中第 $x$ 行第 $y$ 列神经元与第 $l$ 层相连的神经元坐标的集合,$w$ 表示两个神经元相应的连接权值。

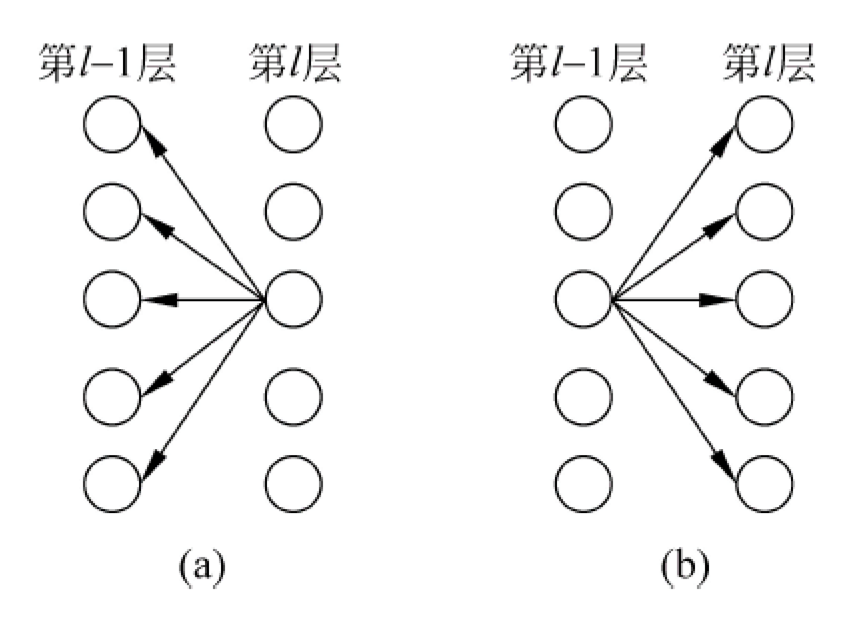



图 8.14 误差反向传播的两种方式

卷积层误差向前层子采样层反向传播的串行计算过程如下。

```
for 卷积层前层各神经元
    确定该神经元与卷积层哪些神经元相连
    利用式(8.22)从卷积层相关神经元“拉”取误差
end for
```

卷积层第 $k$ 个特征图与前层第 $t$ 个特征图相连的权值矩阵中第 $r$ 行第 $y$ 列权值的局部梯度改变量可由式(8.23)计算。

$$\Delta w_{(t,c)}^{(k,t)}=\sum_{x=r}^{fh-kh+r}\sum_{y=c}^{fw-kw+c}d\left(\mathrm{O}_{(x,y)}^{(l,k)}\right)\mathrm{O}_{(r+x,x+y)}^{(l-1,t)} \tag{8.23}$$

卷积层一个特征图上的所有神经元共享一个偏置,与子采样层一样,计算方法也相同。需要解决的问题复杂化和对卷积神经网络性能要求的提高,导致需要的训练数据越来越完备,数据量越来越多,相应地,就需要学习能力更强的网络,而学习能力强的网络需要更多的可训练参数。例如著名的 ImageNet 数据集包含 14 197 122 幅有标签的高清图片[341],文献[388]用于 ImageNet 数据集分类的卷积神经网络包含 650 000 个神经元,60 000 000 个可训练参数。大量的训练数据和学习参数导致计算量大幅增加,串行算法往往需要几个月的训练时间[148]。因此人们研究并行化卷积神经网络。卷积神经网络的训练算法至少存在五种并行方式,即训练样本的并行性、前反向层间的并行性、同层特征图的并行性,特征图中神经元的并行性和神经元权值的并行性[657]。

## 8.9　学习计算理论

学习计算理论主要研究学习算法的样本复杂性和计算复杂性。对于建立机器学习科学，学习计算理论非常重要，否则无法识别学习算法的应用范围，也无法分析不同方法的可学习性。收敛性、可行性和近似性是本质问题，它们要求学习的计算理论给出一种令人满意的学习框架，包括合理的约束。这方面的早期成果主要是基于哥尔德(E M Gold)框架。在形式语言学习的上下文中，哥尔德引入收敛的概念，有效地处理了从实例学习的问题。学习算法允许提出许多假设，无须知道什么时候它是正确的，只要确认某个点它的计算是正确的假设。由于哥尔德算法的复杂性很高，因此这种风范并没有在实际学习中得到应用。

基于哥尔德学习框架，萨皮罗提出了模型推理算法研究形式语言与其解释之间的关系，也就是形式语言的语法与语义之间的关系。模型论把形式语言中的公式、句子理论和它们的解释-模型，当作数学对象进行研究。萨皮罗(E Y Shapiro)模型推理算法只要输入有限的事实就可以得到一种理论输出[672]。

1984 年瓦伦特(L G Valiant)提出一种新的学习框架[774]。它仅要求与目标概念具有高概率的近似，而并不要求目标概念精确的辨识。豪斯勒(Haussler)应用 Valiant 框架分析了变型空间和归纳偏置问题，并给出了样本复杂性的计算公式。

### 8.9.1　哥尔德学习理论

哥尔德语言学习理论研究引入两个基本概念，即极限辨识和枚举辨识，这对早期的归纳推理的理论研究起了非常重要的作用[274]。极限辨识把归纳推理看作一种无限过程，归纳推理方法的最终或极限行为可以看作是它的成功标准。假设 $M$ 是一种归纳推理方法，它企图正确地描述未知规则 $R$。假设 $M$ 重复运行，$R$ 的实例集合则越来越大，形成 $M$ 推测的无限序列 $g_1$，$g_2$，…。如果存在某个数 $m$，使得 $g_m$ 是 $R$ 的正确描述：

$$g_m = g_{m+1} = g_{m+2} = \cdots \tag{8.24}$$

那么 $M$ 在这个实例序列的极限正确地辨识 $R$。$M$ 可以看作对未知规则 $R$ 学习越来越多，成功地修改它关于 $R$ 的推测。如果有限次后 $M$ 停止修改它的推测，最后的推测就是 $R$ 的正确描述，那么在这个实例序列的极限 $M$ 正确地辨识 $R$。注意，$M$ 不能确定它是否会收敛到一个正确的假设，因为新的数据与当前的推测是否会发生矛盾并不知道。

枚举辨识是第一种方法推测多项式序列的抽象，即对可能的规则空间进行系统搜索，直到发现与迄今为止的所有数据相一致的推测。假设规定了规则的具体领域，有一个描述枚举，即 $d_1$，$d_2$，$d_3$，…以至于领域中的每一条规则在枚举中有一种或多种描述。给定一条规则的某个实例集合，枚举辨识方法将通过这个表，找到第一个描述 $d_1$，即与给定的实例相容，那么推测为 $d_1$。这种方法不能确定是否会达到正确的极限辨识。如果实例表示和相容关系满足下面两个条件，那么枚举方法保证极限辨识该领域中的全部规则。

(1) 一个正确假设总是与给定的实例相容。

(2) 任何不正确的假设与实例足够大的集合或与全部集合不相容。

为了枚举方法是可计算的，枚举 $d_1$，$d_2$，$d_3$，…必须是可计算的，它必须能够计算给定的

描述与给定的实例集合是相容的。

**算法 8.2** 枚举辨识算法

输入：

- 一组表达式的集合 $E = e_1, e_2, \cdots$。
- 谕示(Oracle) TE 提供足够的目标实例集。
- 排序信息的谕示 LE。

输出：

一系列假设断言 $H_1$，$H_2$，…，每个假设 $H_i$ 都在 $E$ 中，并与第 $i$ 个实例一致。

过程：

```
1. 初始化, i ←1;
2. examples ←mptyset;
3. Loop:
      3.1 调用 TE(), 将 example 加到集合 examples;
      3.2 While LE(e_i, + x) = no, 对正例集 + x, 或者
               LE(e_i, - x) = yes, 对反例集 - x,
        i ←i + 1;
4. 输出 e_i.
```

## 8.9.2 模型推理系统

模型推理问题是科学家所面临的问题抽象，他们在具有固定概念框架的某种领域里工作，进行试验，试图找到一种理论可以解释他们的结果。在这种抽象中研究的领域是对给定的一阶语言 $L$ 某种未知模型 $M$ 的领域，实验是检测 $M$ 中 $L$ 语句的真值，目标是寻找一组正确假设，它们包含全部正确的可测试的句子。

$L$ 语句分成两个子集：观测语言 $L_o$ 和假设语言 $L_h$。假设

$$\square \in L_o \subset L_h \subset L'$$

其中□是空语句。那么模型推理问题可以定义如下：假设给定一阶语言 $L$ 两个子集：观测语言 $L_o$ 和假设语言 $L_h$。另外对 $L$ 的未知模型 $M$ 给定一种处理机制 Oracle。模型推理问题是寻找 $M$ 的一种有限的 $L_o$——完备公理化。

求解模型推理问题的算法称为模型推理算法。模型 $M$ 的枚举是一个无限序列 $F_1$，$F_2$，$F_3$，…，其中 $F_i$ 是关于 $M$ 的事实，$L_o$ 的每个语句 $\alpha$ 发生在事实 $F_i = <\alpha, V>$，$i>0$。模型推理算法一次读入给定观测语言 $L_o$ 的模型的一种枚举，一个事实，产生假设语言 $L_h$ 的语句的有限集称为算法的推测。一种枚举模型推理算法如下。

**算法 8.3** 枚举模型推理算法

```
h 是整个递归函数.
设 S_false 为{□}, S_true 为{ }, k 为 0.
repeat
    读入下一个事实 F_n = < α, V >
    α 加到 Sv
    while 有一个 α∈ S_false 以至于 T_k ├_n α
          或有一个 α_i ∈ S_true 以至于 T_k ¬ ├_n(i) α_i do
          k = k + 1
```

输出 $T_k$.

forever

上面算法中 $T \vdash_n \alpha$ 表示在推导 $n$ 步或少于 $n$ 步时，假设语句 $T$ 可以推导出 α。$T \dashv\vdash_{n(i)} \alpha_i$ 表示在推导 $n$ 步或少于 $n$ 步时，假设语句 $T$ 不能推出 $\alpha$。推导中假设是单调的。萨皮罗(E Y Shapiro)证明这种算法是极限辨识。这种算法功能强且灵活，可以从事实推出理论，是一种递增算法。

### 8.9.3　大概近似正确学习理论

瓦伦特认为一个学习机必须具备下列性质。

(1) 机器能够证明地学习所有类的概念。更进一步，这些类可以特征化。

(2) 对于通用知识概念类是合适的和不平常的。

(3) 机器演绎所希望的程序的计算过程要求在可行的步数内。

学习机由学习协议和演绎过程组成。学习协议规定从外部获得信息的方法。演绎过程是一种机制，学习概念的正确识别算法是演绎的。从广义来看，研究学习的方法是规定一种可能的学习协议，使用这种协议研究概念类，识别程序可以在多项式时间内演绎。具体协议允许提供两类信息。第一种是学习者对典型数据的访问，这些典型数据是概念的正例。要确切地说，假设这些正例本质上有一种任意确定的概率分布。调用子程序 EXAMPLES 产生一种这样的正例。产生不同例子的相对概率是分布确定的。第二个可用的信息源是 ORACLE。在最基本的版本中，当提交数据时，它将告诉学习该数据是否是概念的正例示。

假设 $X$ 是实例空间，一个概念是 $X$ 的一个子集。如果实例在概念中则为正例，否则为反例。概念表示是一种概念的描述，概念类是一组概念表示。学习模型是概念类的有效的可学习性。大概近似正确 PAC(Probably Approximately Correct )学习理论仅要求对目标概念的近似具有极高的概率。允许学习者产生的概念描述与目标概念有一个小的偏差 ε，它是学习算法的一个输入参数。并且，允许学习者失败的概率为 $\delta$，这也是一个输入参数。两种概念之间的差别采用在实例空间 $X$ 的分布概率 $D$ 来评测：

$$\mathrm{diff}_D(c_1, c_2) = \sum_{x \in X, c_1(x) \neq c_2(x)} D(x) \tag{8.25}$$

根据协议，一个概念类 $C$ 是可学习的当且仅当有一种算法 A，使用协议，对所有的目标概念表示 $c^* \in C$ 和全部分布 $D$，

(1) 执行时间是与 $\frac{1}{\varepsilon}, \frac{1}{\delta}, c^*$ 数目和其他相关参数有关的多项式。

(2) 输出 $C$ 中的概念 $c$ 具有概率 $1-\delta$，

$$\mathrm{diff}_D(c, c^*) < \varepsilon \tag{8.26}$$

PAC 学习理论中，有两种学习复杂性测度。一种是样本复杂性。这是随机实例的数目，用以产生具有高的概率和小的误差。第二种性能测度是计算复杂性，定义为最坏情况下以给定数目的样本产生假设所要求的计算时间。PAC 学习理论，仅要求学习算法产生的假设能以高的概率很好接近目标概念，并不要求精确地辨识目标概念。这种学习理论是“大概近似正确”辨识。PAC 学习理论比 Gold 学习理论更有实际意义。

# >>第 9 章
# 记忆

记忆是人脑对经历过事物的识记、保持、再现或再认，它是进行思维、想像等高级心理活动的基础。由于记忆，人才能保持过去的反映，使当前的反映在以前反映的基础上进行，使反映更全面、更深入。有了记忆，人才能积累经验，扩大经验。记忆是心理在时间上的持续。有了记忆，先后的经验才能联系起来，使心理活动成为一个发展的过程，使一个人的心理活动成为统一的过程，并形成他的心理特征。记忆是反映机能的一个基本方面。

## 9.1　概述

记忆是在人脑中积累、保存和提取个体经验的心理过程。运用信息加工的术语，就是人脑对外界输入的信息进行编码、存储和提取的过程。人们感知过的事物，思考过的问题，体验过的情感和从事过的活动，都会在头脑中留下不同程度的印象，这个就是记的过程；在一定的条件下，根据需要这些储存在头脑中的印象又可以被唤起，参与当前的活动，得到再次应用，这就是忆的过程。从向脑内存储到再次提取出来应用，这个完整的过程总称为记忆。

记忆包括三个基本过程：信息进入记忆系统——编码，信息在记忆中储存——保持，信息从记忆中提取出来——提取。编码是记忆的第一个基本过程，它把来自感官的信息变成记忆系统能够接收和使用的形式。一般说来，我们通过各种感觉器官获取的外界信息，首先要转换成各种不同的记忆代码，即形成客观物理刺激的心理表征。编码过程需要注意的参与。注意使编码有不同的加工水平，或采取不同的表现形式。例如对于一个汉字，你可以注意它的字形结构、字的发音或字的含义，形成视觉代码、声音代码或语义代码。编码的强弱直接影响着记忆的长短。当然，强烈的情绪体验也会加强记忆效果。总之，如何对信息编码直接影响到记忆的储存和以后的提取。一般情况下，对信息采用多种方式编码会收到更好的记忆效果。

已经编码的信息必须在头脑中得到保存，在一定时间后才可能被提取。但信息的保存并不都是自动的，在大多数情况下，为了日后的应用，我们必须想办法努力将信息保存下来。已经储存的信息还可能受到破坏，出现遗忘。心理学家研究记忆主要关心的就是影响记忆储存的因素，以便与遗忘做斗争。

保存在记忆中的信息，只有在被提取出来加以应用时，才有意义的。提取有两种表现方式：回忆和再认。日常所说“记得”指的就是回忆。再认较容易，原因是原刺激呈现在眼前，你有各种线索可以利用，需要的只是确定它的熟悉程度。一些学习过的材料无法回忆或者再认出来，它们是否在头脑里完全消失了呢？不是的。记忆痕迹并不会完全消失，用再学习可以很好地证明这一点。即让被试先后两次学习同一材料，每次达到同样的熟练水平，再次

学习所需要的练习次数或时间必定要少于初次学习，两次所用时间或次数之差就表示了保存的数量。

根据记忆的内容，可以把记忆分成如下四种。

（1）形象记忆。以感知过的事物形象为内容的记忆叫做形象记忆。这些具体形象可以是视觉的，也可以是听觉的、嗅觉的、触觉的或味觉的形象，如人们对看过的一幅画，听过的一首乐曲的记忆就是形象记忆。这类记忆的显著特点是保存事物的感性特征，具有典型的直观性。

（2）情绪记忆。以过去体验过的情绪或情感为内容的记忆。如学生对接到大学录取通知书时的愉快心情的记忆等。人们在认识事物或与人交往的过程中，总会带有一定的情绪色彩或情感内容，这些情绪或情感也作为记忆的内容而被存储进大脑，成为人的心理内容的一部分。情绪记忆往往是一次形成而经久不忘的，对人的行为具有较大的影响作用。情绪记忆的印象有时比其他形式的记忆印象更持久，即使人们对引起某种情绪体验的事实早已忘记，但情绪体验仍然保持着。

（3）逻辑记忆。以思想、概念或命题等形式为内容的记忆。如对数学定理、公式、哲学命题等内容的记忆。这类记忆是以抽象逻辑思维为基础的，具有概括性、理解性和逻辑性等特点。

（4）动作记忆。以人们过去的操作性行为为内容的记忆。凡是人们头脑里所保持的做过的动作及动作模式，都属于动作记忆。这类记忆对于人们动作的连贯性、精确性等具有重要意义，是动作技能形成的基础。

以上四种记忆形式既有区别，又紧密联系在一起。如动作记忆中具有鲜明的形象性。逻辑记忆如果没有情绪记忆，其内容是很难长久保持的。

## 9.2　记忆系统

根据记忆操作的时间长短，人类记忆有三种类型：感觉记忆、短时记忆和长时记忆。三者的关系可以由图 9.1 表示出来。来自环境的信息首先到达感觉记忆。如果这些信息被注意，它们则进入短时记忆。正是在短时记忆中，个体把这些信息加以改组和利用并作出反应。为了分析存入短时记忆的信息，你会调出储存在长时记忆中的知识。同时，短时记忆中的信息如果需要保存，也可以经过复述存入长时记忆。在图 9.1 中，箭头表明信息流在三种存储模型中的运行方向。

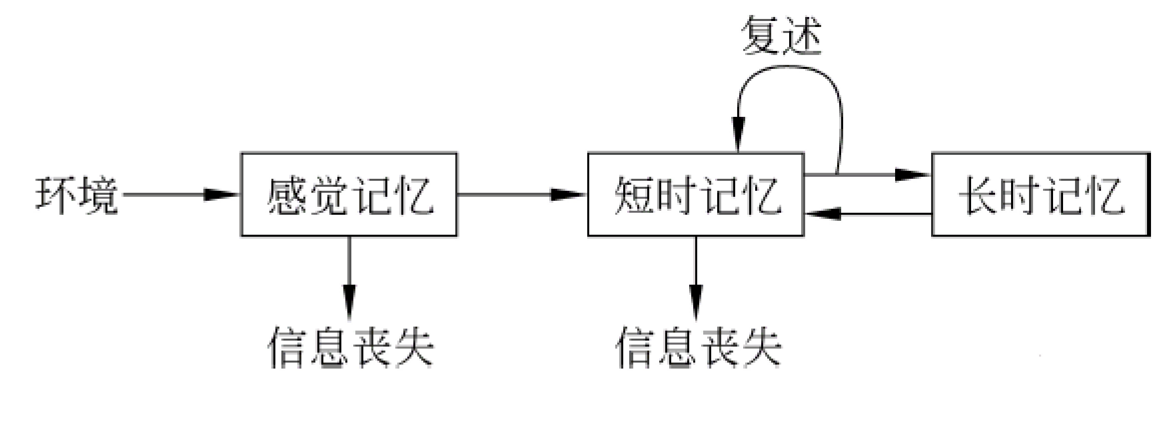



图 9.1　记忆系统

阿特金森（R Atkinson）和雪芙林（R M Shiffrin）在 1968 年对其记忆系统模型进行扩充，扩展的模型如图 9.2 所示。从图 9.2 中可以看出，记忆系统的模型主体由感觉记忆（感

觉登记)、短时记忆(短时储存)和长时记忆(长时储存)三部分构成,所不同的是,他们加入了控制过程这一内容,认为控制过程在三种储存过程都起作用。该模型还有一个值得关注的要点就是它对长时记忆信息的认识。模型认为长时记忆中的信息是不会消失的,其信息是不消退的自寻地址库。

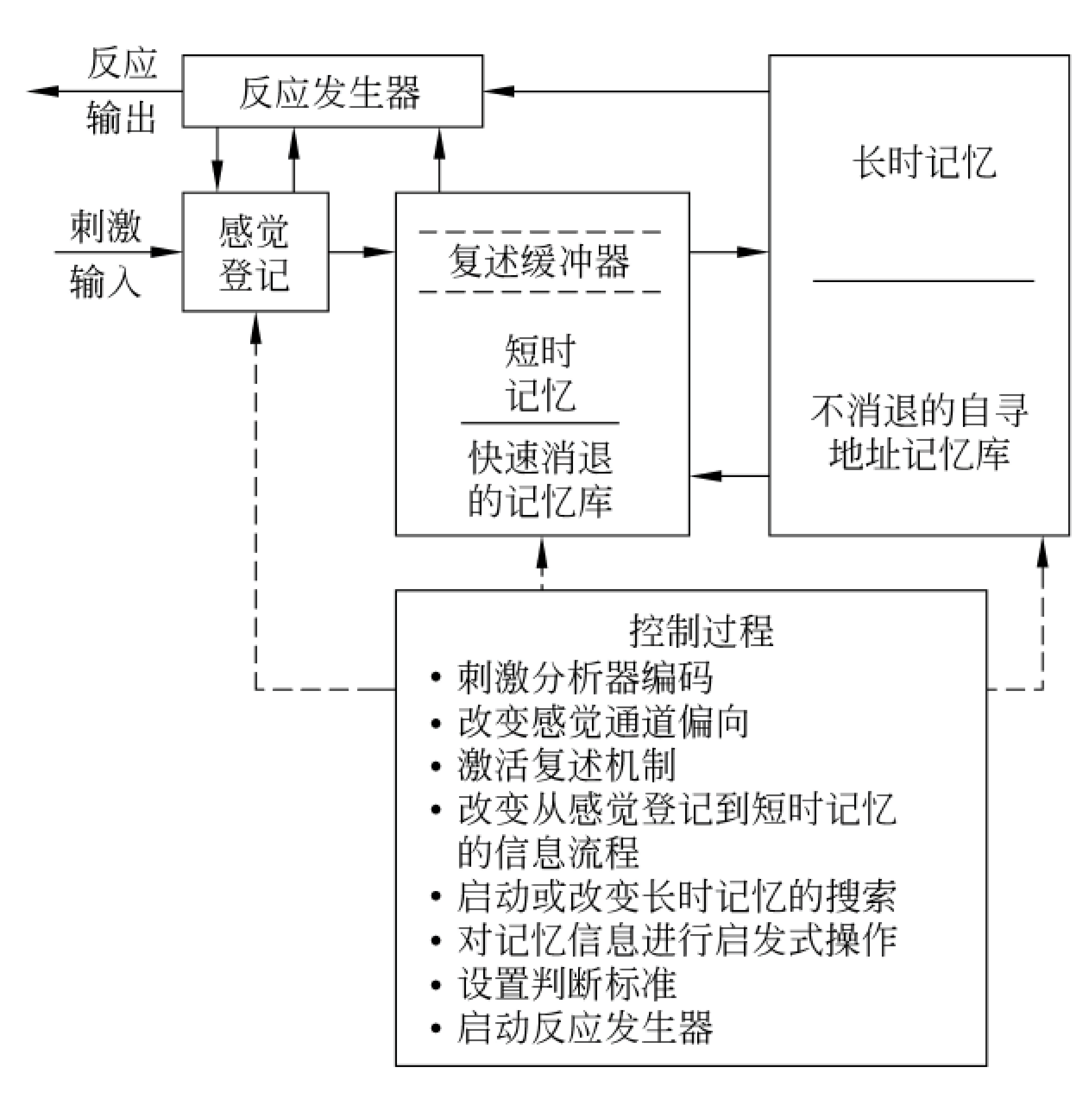



图9.2 记忆系统模型

### 9.2.1 感觉记忆

感觉记忆又称感觉寄存器或瞬时记忆,是感觉信息到达感官的第一次直接印象。感觉寄存器只能将来自各个感官的信息保持几十到几百毫秒。在感觉寄存器中,信息可能受到注意,经过编码获得意义,继续进入下一阶段的加工活动,如果不被注意或编码,它们就会自动消退。

各种感觉信息在感觉寄存器中以其特有的形式继续保存一段时间并起作用,这些存储形式就是视觉表象和声音表象,称视象和声象。表象可以说是最直接、最原始的记忆。表象只能存在很短的时间,如最鲜明的视象也不过持续几十秒钟。感觉记忆具有下列特征:

(1) 记忆非常短暂;

(2) 有能力处理像感受器在解剖学和生理学上所能操纵的同样多的物质刺激能量;

(3) 以相当直接的方式把信息编码。

斯伯林(George Sperling)的研究证实了独立感觉记忆存储的概念。在这些研究中,他利用了一种叫速示器的工具,对被试者以非常短和精确的时间呈现视觉刺激。呈现可以控制到接近1/1000s。当斯伯林以1.5/1000～500/1000s时间呈现符号(字母或数字)时,他注意到不管在一次特定呈现中只呈现少量符号,还是大量符号,他的被试者只能准确报告四五个符号。为了表明被试者所能报告的符号是有限制的,而不是他看到的符号有限制,采用部分报告技术。给被试者呈现12个符号,排成3排,每排4个符号。每个刺激呈现50/1000s,

跟着立即发出三个不同音中的一个。指示被试者如果他们听到第一个音,报告第一排符号;如果听到第二个音,报告第二排符号;如果听到第三个音,报告第三排符号。

采用这个技术,被试者平均能准确地报告任何一排中所要求的字母的 76%,如果他们在报告全显示的 12 个符号时能达到同样的百分比,可以指望大约有准确报告 9 个符号的平均数。但是,不管呈现多少符号,准确报告的符号只有 4 个或 5 个。

为什么有这种不一致?斯伯林提出,被试经验到全显示的视觉映像,这个映像就用作视觉记忆存储。但是这个映像很快消失,如果被试者从这个映象捡取一排报告,当他回到下一排时,映像已经消失。他能够很好地报告他们所捡取的第一排,但是之后就报告得很少。实际上,他看到的符号比能报告得多。

目前关于感觉记忆的研究主要在听觉和视觉通道上进行。视觉的感觉记忆被称为图像记忆(Iconic Memory),听觉的感觉记忆被称为声象记忆(Echoic Memory)。

## 9.2.2 短时记忆

在感觉记忆中经过编码的信息,进入短时记忆后经过进一步的加工,再从这里进入可以长久保存的长时记忆。信息在短时记忆中一般只保持 20～30s,但如果加以复述,便可以继续保存。复述保证了它的延缓消失。短时记忆中储存的是正在使用的信息,在心理活动中具有十分重要的作用。首先,短时记忆扮演着意识的角色,使我们知道自己正在接收什么以及正在做什么。其次,短时记忆使我们能够将许多来自感觉的信息加以整合构成完整的图像。再次,短时记忆在思考和解决问题时起着暂时寄存器的作用。最后,短时记忆保存着当前的策略和意愿。这一切使得我们能够采取各种复杂的行为直至达到最终的目标。正因为发现了短时记忆的这些重要作用,在当前大多数研究中被改称为工作记忆。和感觉记忆中可用的大量信息对比,短时记忆的能力是相当有限的。如果给被试者一个数字串,例如 6—8—3—5—9,他能立即背出来。如果是 7 个以上数字的数字串,一般人就不能很好背出来。1956 年,美国心理学家米勒明确提出,短时记忆容量为 7±2 个组块(Chunk)。组块是指将若干较小单位联合成熟悉的、较大的单位的信息加工,也指这样组成的单位。组块既是过程,也是单位。知识经验与组块:组块的作用在于减少适时记忆中的刺激单位,而增加每一单位所包含的信息。人的知识经验越丰富,组块中所包含的信息越多。与组块相似,但它不是意义分组,各成分之间不存在意义联系。为了能记忆较长的数字串,把数字分组,从而有效地减少数字串中独立成分的数量,是个有效的办法。这种组织称作组块,在长时记忆中发挥巨大作用。

有人曾经指出,刺激信息是根据它的听觉特性存储在短时记忆中的。这就是说,即使是凭视觉接收的信息,将按听觉的声学的特性编码。例如你看到一组字母 B—C—D,你是根据它们的读音[bi:]-[si:]-[di:]编码,而不是根据它们的字形编码。

人类的短时记忆编码也许具有强烈的听觉性质,但也不能排除其他性质的编码。不会说话的猴子,也能够做短时记忆的工作。例如,给它们看过图形的一个样本以后不久,它们会在两个彩色几何图形中挑选出一个。

图 9.3 给出了短时记忆复述缓冲器。短时记忆由若干槽构成。每一个槽相当于一个信息通道。来自感觉记忆的信息单元分别进入不同的槽。缓冲器的复述性加工有选择地将槽中的信息进行复述。被复述的槽中的信息将进入长时记忆中。而没有被复述的槽中的信息

将被清除出短时储存区而丧失。

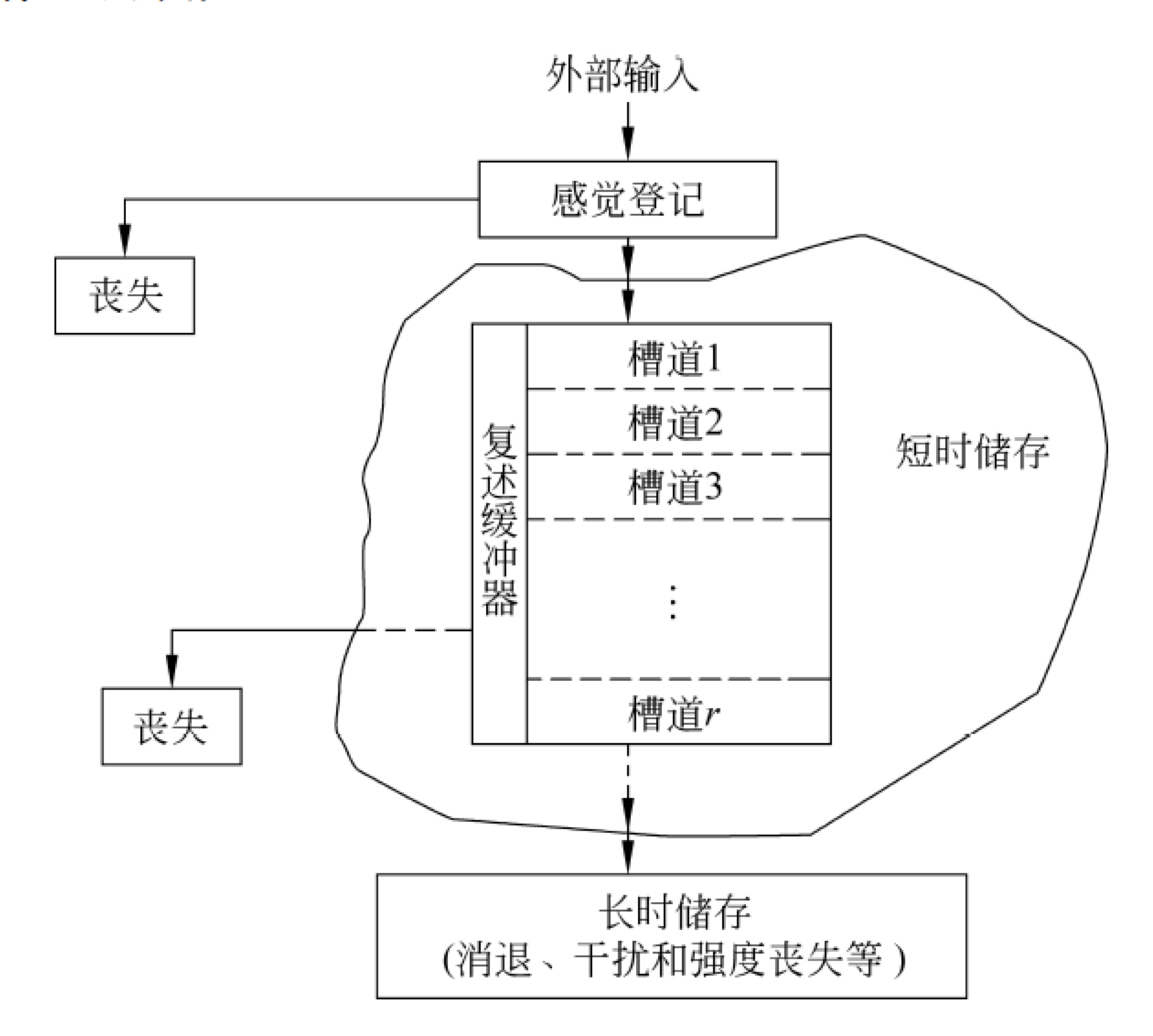



图9.3 短时记忆复述缓冲器

各槽中的信息保持的时间是不一样的。信息在槽中保持的时间越长,越有可能进入长时记忆中,也越有可能被来自感觉记忆的新的信息冲挤掉。相对而言,长时记忆才是一个真正的信息储存库,但其中的信息也有可能因消退、干扰和强度丧失等原因而产生遗忘。

短时记忆信息提取过程是相当复杂的。它涉及许多问题,并且引出不同的假说,迄今没有一致的看法。

### 1. 斯特恩伯格的经典研究

斯特恩伯格(Saul Sternberg)的研究表明,短时记忆中信息的提取是通过系列扫描,即从头至尾扫描方式来实现的,可以将之理解为扫描模型。

斯特恩伯格的实验可以说是一个经典的研究范式。其实验假设为,如果被试者要对短时记忆集中所有识记项目进行全部扫描后才能对测试项目进行"是"或"否"判断,那么被试者进行正确判断所需的反应时间不应随记忆集的大小而变化(图9.4(a)所示)。而实验结果却如图9.4(b)所示,被试的反应时间会随短时记忆集的增大而延长。这说明,短时记忆的扫描不是进行全部的扫描,而是进行有序的系列扫描。

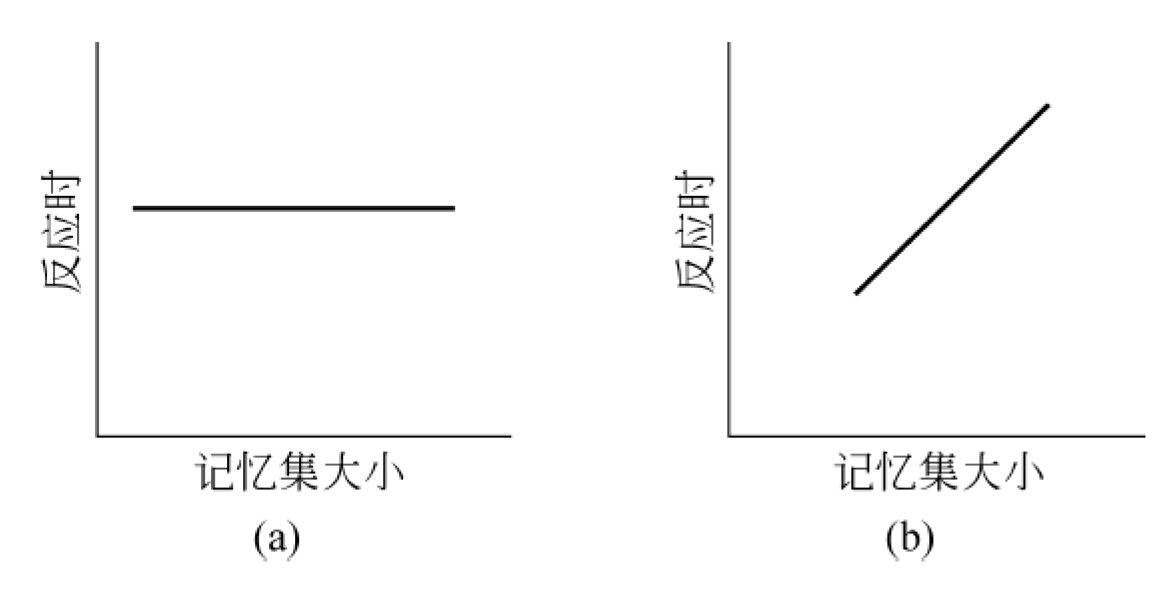



图9.4 斯特恩伯格的扫描试验

斯特恩伯格的理论必须要解决的另一个问题是如果说短时记忆中信息的提取是通过系列扫描而不是平行扫描实现的，那么，这种扫描是从什么地方开始，又是怎样结束的。他认为，自顾不暇时记忆信息提取是以从头至尾的系列扫描方式进行的。同时，判断过程包括比较过程和决策过程两个阶段，因此，进行判断时也不进行自我停止扫描。

## 2. 直接存取模型

维克勒格林(Wickelgren)认为短时记忆中的各个项目不是通过比较来提取的，人可直接通往所要提取的项目在短时记忆中的位置，进行直接提取。

直接存取模型(Direct Access Model)认为，短时记忆中信息的提取并不是通过扫描的方式进行的，大脑可以直接存取所要提取的项目在短时记忆中的位置，进行直接提取。该模型认为，短时记忆中的每一个项目都有一定的熟悉值或痕迹强度，可以据此作出某种判定。在大脑内部有一个判断标准，当熟悉值高于这一标准，则作出"是"反应，低于这一标准则作出"否"反应。熟悉值与标准的偏离程度越高，作出是或否反应的速度也越快。

直接存取模型可以解除系列位置效应(首因效应和近因效应)，但是短时记忆是如何知道识记项目的位置的？如果信息的提取是直接存取的，那么为什么反应时间会随着识记项目的增加而呈线性增加？

## 3. 双重模型

阿特金森(R Atkinson)和鸠拉(J Juola)认为，短时记忆过程中信息的提取既包含扫描方式，也存在直通方式，简言之就是两头直通，中间扫描(见图 9.5)。

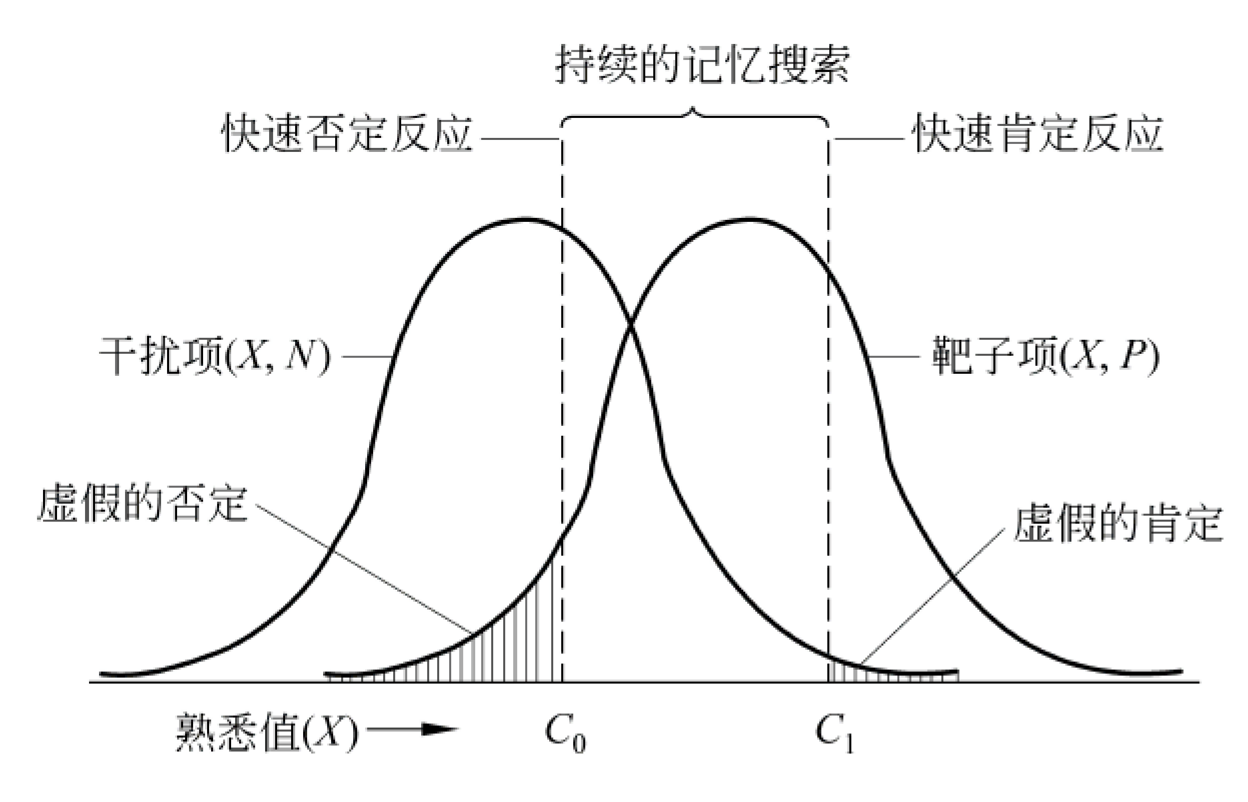



图 9.5　双重模型示意图

由于搜索模型和直接存取模型都有其合理的一面，同时又都有不足，因此有人企图将两者结合起来。阿特金森和鸠拉提出的短时记忆信息提取双重模型就是一个尝试[37]。他们设想，输入的每个字词可按其知觉维量来编码，称为知觉代码；字词还有意义，即有概念代码。知觉代码和概念代码共同构成一个概念结。每个概念结有不同的激活水平或熟悉值。

在大脑内部有两个判定标准；一个是"高标准"(C1)，如果某一探测词的熟悉值达到或高于这个标准，人便可迅速地作出"是"反应；另一个是低标准(C0)，如果某一探测词的熟悉值达到或低于这个标准，人就可迅速地作出"否"反应。照阿特金森和鸠拉看来，这是一个直

接存取过程。但是,对于一个熟悉值低于“高标准”而高于“低标准”的探测词,则要进行系列搜索,才能作出反应,所需反应时间也较多。

短时记忆的研究还发现,信息的加工速率与材料性质或信息类型有一定关系,即加工速率随着记忆容量的增大而提高,容量愈大的材料,扫描也愈快。

已有的实验结果表明,短时记忆信息提取的加工速率与材料性质或信息类型有一定关系。1972 年卡凡诺夫(Cavanaugh)通过统计不同的研究对某类材料的平均实验结果,得出扫描一个项的平均时间,并与相应的短时记忆容量(广度)加以对照,见表 9.1。从表中可以看出一个有趣的现象;加工速率随着记忆容量的增大而提高,容量越大的材料,扫描也越快。现在还难以清楚地解释这个现象。曾经设想,在短时记忆中,信息是以特征来表征的,而短时记忆的储存空间有限,则每个刺激的平均特征数量越大,那么短时记忆能够储存的刺激数量就越小。卡凡诺夫进而认为,每个刺激的加工时间与其平均特征数量成正比,平均特征数量大的刺激需要的加工时间多,反之需要的加工时间则少。这种解释还存在不少疑点。但它却把短时记忆的信息提取、记忆容量和信息表征都联系起来。这确实是一个重要的问题。加工速率反映加工过程的特点,在不同材料的加工速率差别的背后,可能由于记忆容量乃至信息表征等因素的作用而存在着不同的信息提取过程。

**表 9.1　不同类别材料的加工速率与记忆容量**

| | 加工速度(ms) | 记忆容量(项) |
|---|---|---|
| 数　　字 | 33.4 | 7.70 |
| 颜　　色 | 38.0 | 7.10 |
| 字　　母 | 40.2 | 6.35 |
| 字　　词 | 47.0 | 5.50 |
| 几何图形 | 50.0 | 5.30 |
| 随机图形 | 68.0 | 3.80 |
| 无意义音节 | 73.0 | 3.40 |

短时记忆中的信息遗忘,通过干扰作业法(Peterson-Peterson 方法)发现如下规律。

(1) 短时记忆中信息可以保持 15～30s。

(2) 如果得不到复述,那么短时记忆中的信息将会迅速遗忘,如图 9.6 所示。

(3) 只要短时记忆识记项的数量不变,识记材料性质的改变对短时记忆的遗忘没有什么大的影响。

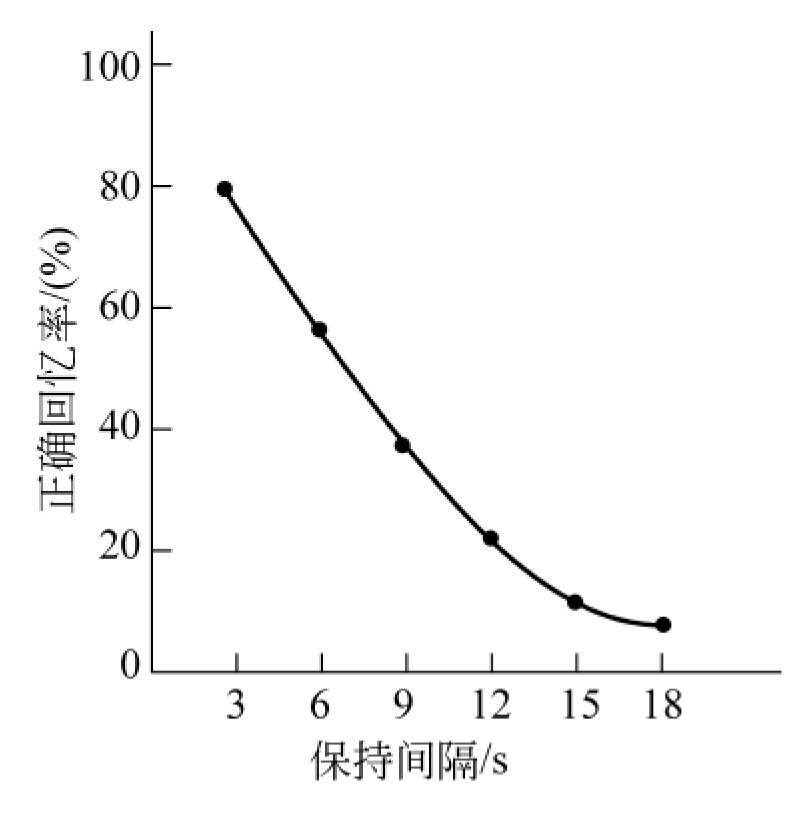



图 9.6　阻止复述后的短时记忆的遗忘速率

### 9.2.3　长时记忆

长时记忆是指保持时间在一分钟以上的信息存储。长时记忆的能力，是一切记忆系统中最大的一个。关于长时记忆的容量、存储、恢复及持续时间，都要用实验说明。一个东西记住以后能持续多长时间，测量的结果是不定的。因为注意不稳定，保持时间就短；如果加以复述，保持时间就可以很长。长时记忆的的容量是无限的。每个组块的存入时间需要8s。长时记忆里的东西要先转入短时记忆，然后才能恢复和应用。长时记忆的恢复，第一个数字用 2s，以后每个数字用 200～300s。我们可以用不同位数的数字来做实验，如使用34 597 743 218 三个数目，测量恢复不同位数的数字所需要的时间。实验结果表明，恢复两位数字用 2200ms，三位数字用 2400ms，六位数字用 3000ms。

蔡卡尼克(Bluma Zeigarnik)效应是让被试者把一些工作做完，另一些工作没做完就中断。过一段时间让他回忆都做过哪些工作。结果，他回忆未完成的工作比回忆已完成的工作要好些。这说明有些活动还没做完就搁下来时，在另一空间还在继续进行活动。在长时记忆里，有些东西比另一些东西容易被提取，因为它们的阈限低。有些阈限高，需要多一些线索才能提取出来。没有完成的工作阈限很低，容易激活和扩散。扩散随网络进行，达到该事件所在的位置时就提取出来了。

很多心理学家已经提出感觉记忆和短时记忆的迅速的、被动的消失，但是很少心理学家赞成长时记忆有这样简单的衰退机制，因为很难解释为什么有些材料比另一些材料被遗忘得快些？遗忘是否与原来的学习材料的完善程度有关？遗忘是否受学习材料时间和回忆材料时间之间发生的事情的影响？研究这些问题的很多心理学家相信，长时记忆的消失是由于干扰。这是一种被动的观点。有一些关于遗忘的观点提出一个比较主动的过程，作为干扰的一种补充或替代。弗洛伊德提出由于压抑而遗忘的观点。如果记住一种材料在心理上会是极其痛苦和有威胁的，那么，这种材料就难于回忆。另一个主动遗忘观点来自巴特莱特(Bartlett)的“创见性的遗忘”。当没有得到精确的记忆时，模仿创造一点与记忆相像的东西，你就接近这个记忆。表 9.2 给出了记忆系统中三种不同类型的记忆的特点。

**表 9.2　三种记忆类型的比较**

| 记忆系统 | 操作的时间间隔 | 能　力 | 组织或编码类型 | 遗忘机制 |
|---|---|---|---|---|
| 感觉记忆 | 一秒的几分之几 | 只限于感受器所能记下的多少 | 物质刺激的相当直接的后像 | 消极的衰退 |
| 短时记忆 | 少于 1min | 只有少数项(5～9个) | 间接的编码包括大量听觉组织 | 消极的衰退 |
| 长时记忆 | 1min 以上到许多年 | 很大，几乎无限 | 很复杂的编码 | 干扰和忘却，压抑，创建性的遗憾 |

人类的记忆系统与计算机的存储系统极其相似。在计算机系统中存储层次可分为高速缓冲存储器、主存储器、辅助存储器三级，构成速度由快到慢、容量由小到大的多级层次存储器，以最优的控制调度算法和合理的成本，构成其有可接受性能的存储系统。

# 9.3 长时记忆

长时记忆的存储形式是指信息在人脑中的内部表示。这方面的研究有一定的困难,因为信息存储的内部结构不能为人所观察,只能通过间接的办法进行研究。我们可以用计算机模拟的方法来研究信息的内部结构,逐渐加深对人类记忆的认识。

## 9.3.1 长时记忆的类型

人类的记忆可以分为程序性记忆和命题记忆。程序性记忆是保持有关操作的技能,主要由知觉运动技能和认知技能组成。命题记忆是存储用符号表示的知识,反映事物的实质。程序性记忆和命题记忆都是反映某个人现在的经验和行动受到以前的经验和行动的影响的记忆,这一点是相同的。同时,它们之间又有区别。第一,程序性记忆中表示的方法只有一种,要进行技能的研究;命题知识的表示可以各种各样,与行动完全不同。第二,关于知识的真假问题,熟练的过程没有真假之分;真假问题只是出现在对世界的认识以及自身与世界的关系的知识方面。第三,两种信息习得的形式不同。过程信息必须通过一定的练习,而命题信息只要一次机会的学习。第四,熟练的行动是"自动"执行的,命题信息的表达要给以注意。

命题记忆更进一步分为情景记忆和语义记忆。前者是存储个人发生的事件和经验的记忆形式。后者是存储个人理解的事件的本质的知识,即记忆关于世界的知识。两种记忆的差别如表 9.3 所示。

**表 9.3 情景记忆和语义记忆的比较**

| 区分特性 | 情景记忆 | 语义记忆 |
|---|---|---|
| 信息 | | |
| 输入源 | 感觉 | 理解 |
| 单位 | 事件、情景 | 事实、观念概念 |
| 体制化 | 时间的 | 概念的 |
| 参照 | 自己 | 万物(世界) |
| 真实性 | 个人的信念 | 社会的一致 |
| 操作 | | |
| 记忆内容 | 经验的 | 符号 |
| 时间的符号化 | 有,直接的 | 无,间接的 |
| 感情 | 较重要 | 并不那么重要 |
| 推理能力 | 小 | 大 |
| 文脉依存性 | 大 | 小 |
| 被干涉性 | 大 | 小 |
| 存取 | 按意图 | 自动的 |
| 检索方法 | 按时间或场所 | 按对象 |
| 检索结果 | 记忆结构变化 | 记忆结构不变 |
| 检索原理 | 协调的 | 开放式 |
| 想起内容 | 被记忆的过去 | 被表示的知识 |
| 检索报告 | 觉得 | 知道 |
| 发展顺序 | 慢 | 快 |
| 小儿健忘症 | 受到障碍 | 不受障碍 |

续表

| 区 分 特 性 | 情 景 记 忆 | 语 义 记 忆 |
|---|---|---|
| 应用 | | |
| 教育 | 无关系 | 有关系 |
| 通用性 | 小 | 大 |
| 人工智能 | 不清理 | 非常好 |
| 人类智能 | 无关系 | 有关系 |
| 经验证据 | 忘却 | 言语分析 |
| 实验室课题 | 特定的情景 | 一般的知识 |
| 法律证词 | 可以，目击者 | 不行，鉴定人 |
| 记忆丧失 | 有关系 | 无关系 |
| 杰恩斯(J Jaynes)的二分心理 | 无 | 有 |

情景记忆(Episodic Memory)是加拿大心理学家图尔文(Endel Tulving)提出来的。1983 年出版了图尔文的专著 *Elements of Episodic Memory*[770]，专门讨论情景记忆的原理。

情景记忆的基本单位是个人的回忆行为。这种回忆行为是开始于事件或情景生成的经验的主观再现(想起经验)，或者变换到保持信息的其他形式，或者采用它们两者的结合。关于回忆，有许多构成要素和构成要素间的关系。构成要素分两类，一类是观察可能的事件，另一类是假说的构成概念。这种构成要素是情景记忆的要素。情景记忆的要素可以分成两类，即编码和检索。编码是关于某时某种情况的经验的事件的信息，指出变换到记忆痕迹的过程。检索要素主要与检索方式和检索技术有关。图 9.7 给出了情景记忆的要素和它们的关系。

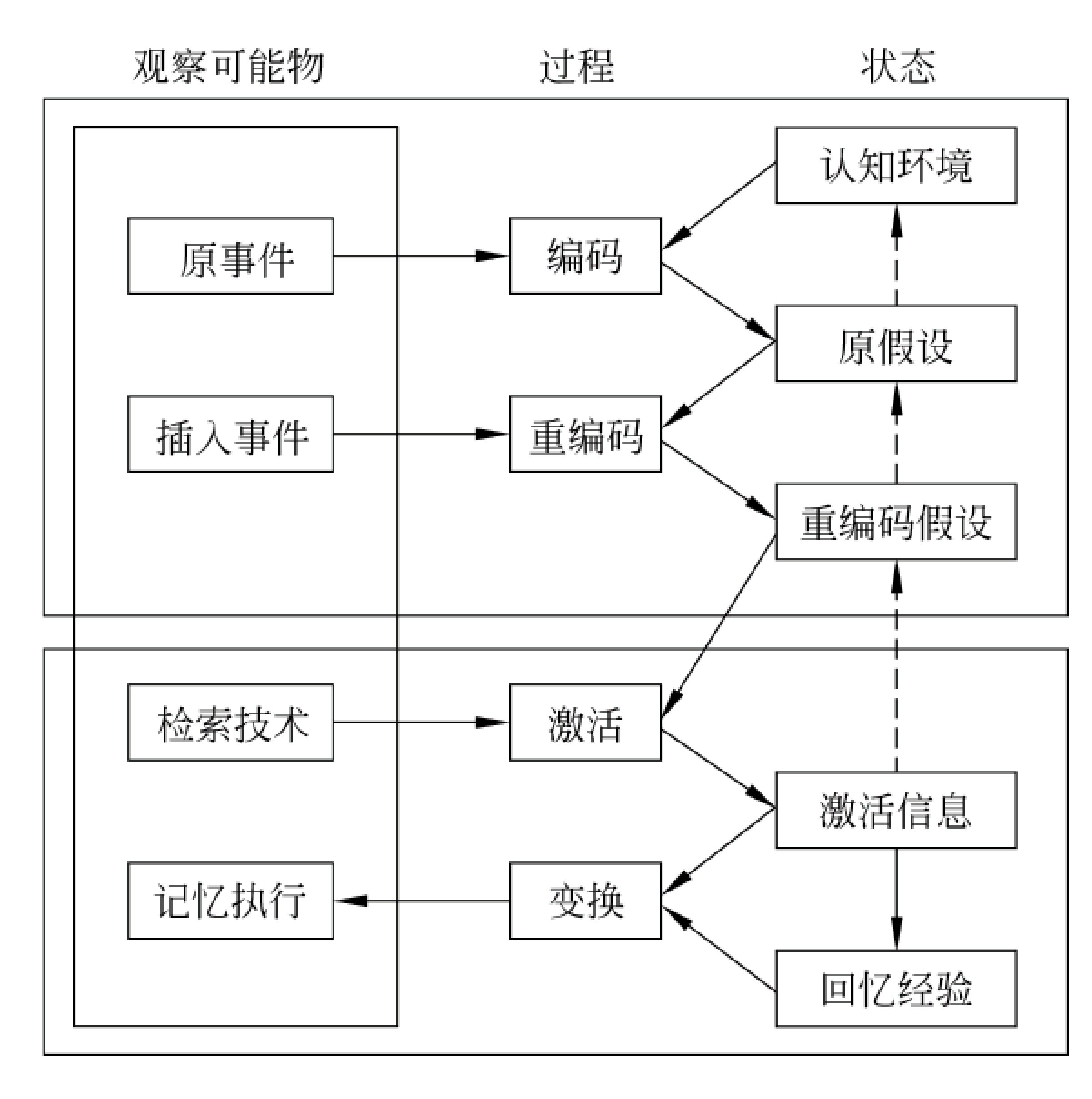



图 9.7　情景记忆的要素和它们的关系

奎连(Quillian)于 1968 年提出的语义记忆是认知心理学中的第一个语义记忆模型[606]。

在认知心理学方面,安得森(Anderson)和鲍威(Bower),鲁梅哈特(Rumelhart)和诺尔曼(Norman)都提出过基于语义网络的各种记忆模型。在这个模型中,语义记忆的基本单元是概念,每个概念具有一定的特征。这些特征实际上也是概念,不过它们是说明另一些概念的。在一个语义网中,信息被表示为一组结点,结点通过一组带标记的弧彼此相连;带标记的弧代表结点间的关系。图 9.8 是一个典型的语义网络。我们用 ISA 链接表示概念结点之间的层次关系,有时还用 ISA 链接把表示具体对象的结点与其相关概念关联起来。ISPART 链接整体与部分的概念结点。例如图 9.8 中,椅子是座位的一部分。

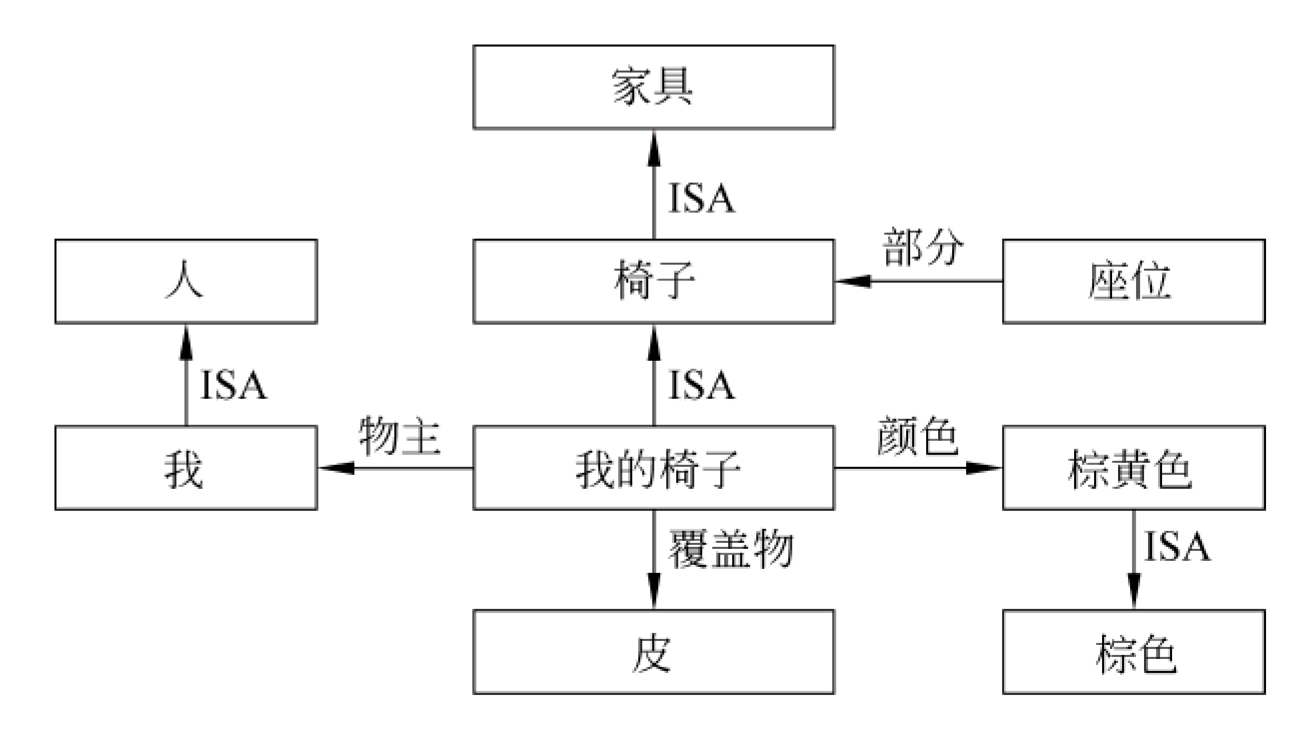



图 9.8 语义网络

从信息编码的角度将长时记忆分为两个系统:即表象系统和言语系统。表象系统以表象代码来储存关于具体的客体和事件的信息。言语系统以言语代码来储存言语信息。两个系统彼此独立又互相联系。因此,人们也把其理论称为两种编码说或双重编码说。

## 9.3.2 长时记忆的模型

### 1. 层次网络模型

奎连等提出了语义记忆的层次网络模型(Hierarchical Network Model)[601]。该模型认为,长时记忆中的基本单元是概念,概念在记忆系统是有联系的,形成一个有层次的结构。图 9.9 小圆圈为结点,代表一个概念;带箭头的连线表示概念之间的从属关系。例如,“鸟”这个概念的上级概念为“动物”,其下级概念为“金丝雀”和“鸵鸟”。连线还表示概念与特征的关系,指明各级概念分别具有的特征,如“鸟”所具有的特征是“有翅膀”“能飞”“有羽毛”。连线把代表各级概念的结点联系起来,并将概念与特征联系起来,构成一个复杂的层次网络。连线在这个网络中实际上是具有一定意义的联想。这个层次网络模型对概念的特征相应的实行分级储存。在每一级概念的水平上,只储存该级同水平的概念,同一级每个概念所具有的共同特征则储存于上一级概念组水平上。图 9.9 是概念体系的一个片断[154],位于最下层的“金丝雀”“鲨鱼”等叫做 0 级概念,“鸟”“鱼”等叫做 1 级概念,“动物”叫做 2 级概念。概念的级别越高越抽象,加工所需要的时间也越长。在每一个级别上,只储存该级概念独有的特征。因此一个概念的意义或内涵由该概念与其他相联的概念的特征来决定。

### 2. 激活扩散模型

激活扩散模型(Spreading Activation Model)是卡林斯(J Collins)等提出的。它也是一

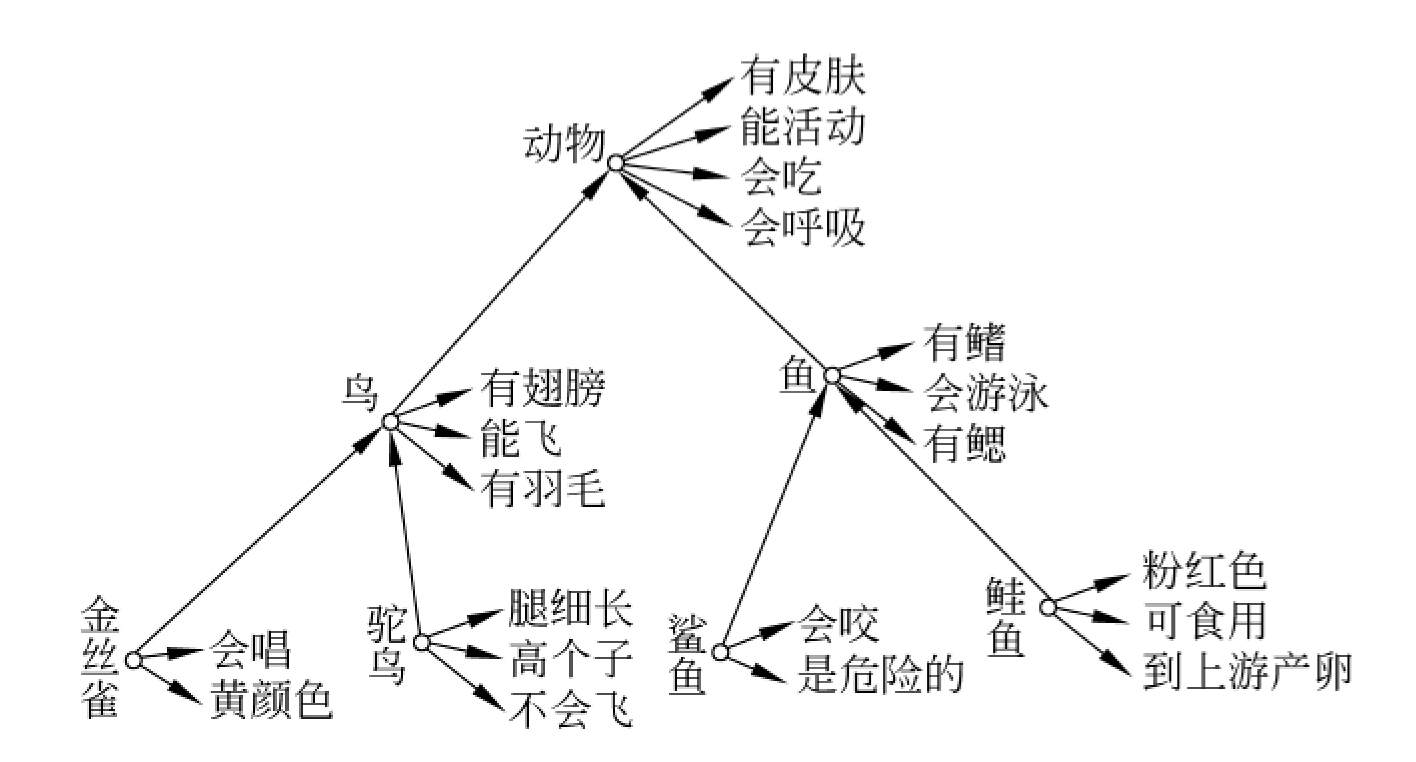



图 9.9　语义记忆的层次网络模型

个网络模型[155]。但与层次网络模型不同，它放弃了概念的层次结构，而以语义联系或语义相似性将概念组织起来，图 9.10 是激活扩散模型的一个片断。图中方框为网络的结点，代表一个概念。概念之间的连线表示它们的联系，连线的长短表示联系的紧密程度，连线越短，表明联系越紧密，两个概念有越多的共同特征；或者两个结点之间通过其共同特征有越多的连线，则两个概念的联系越紧密。

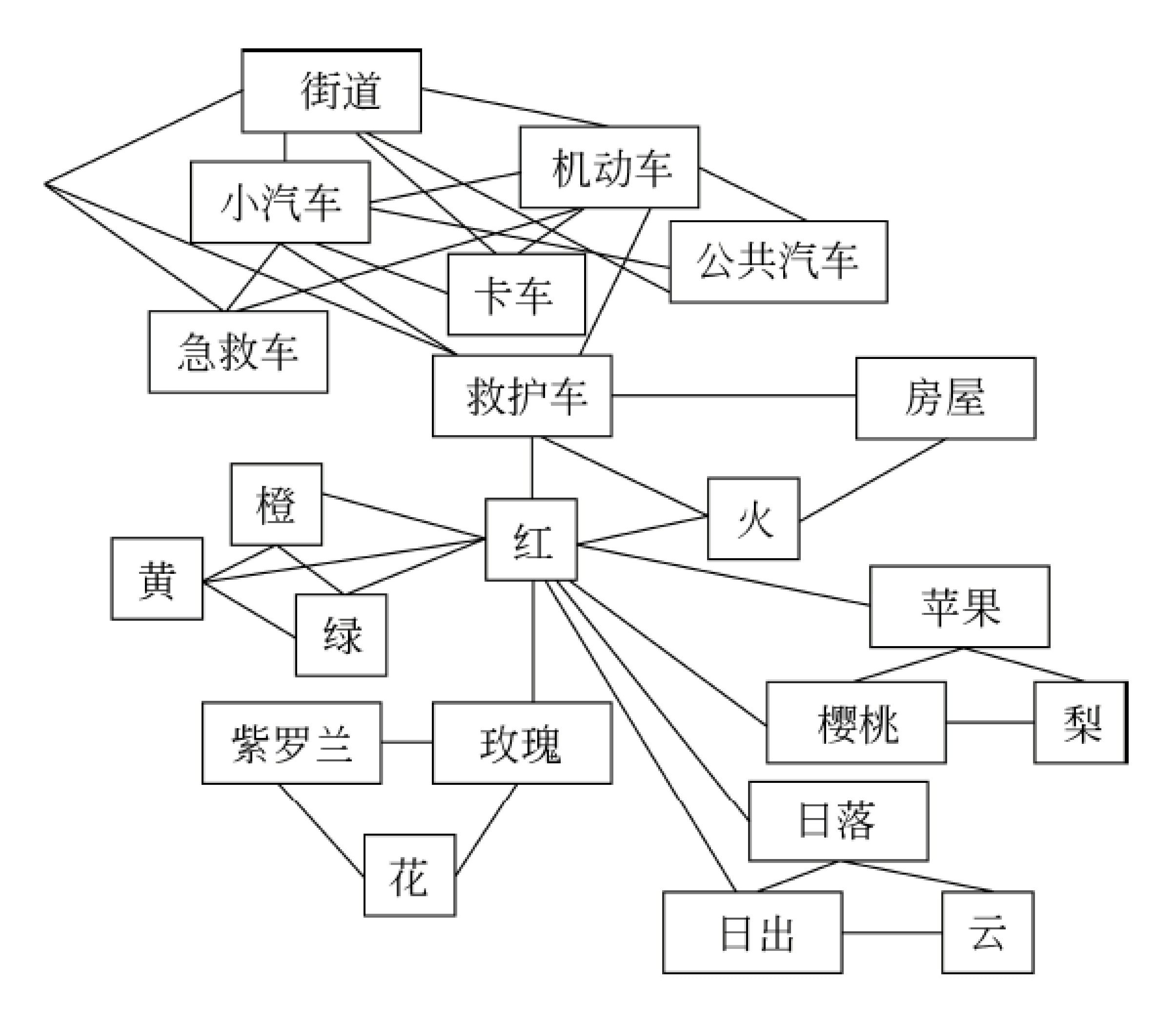



图 9.10　激活扩散模型片断

当一个概念受刺激或被加工，该概念所在的网络节点便被激活，然后激活便沿连线向四周扩散。这种激活的数量是有限的，一个概念越是长时间受到加工，释放激活的时间也越长，从而有可能形成熟悉效应；另一方面，激活也遵循能量递减的规律。该模型是对层次网络模型的修正，它认为诸概念的特征可以在同一层级上也可以不在同一层级上，概念间的联系是激活扩散的方式，它以连线的长短说明范畴大小效应，而且也可以说明其他效应，可以认为它是“人化了的”层次网络模型。

### 3. 集理论模型

集理论模型是梅叶尔(D E Meyer)提出的[497]。在这个模型中(见图 9.11)，基本的语义单元仍为概念。每个概念都由一集(Set)信息或要素来表征。这些信息集可分样例集和属

性集或特征集。样例集是指一个概念的一些样例,如"鸟"概念的样例集包括"知更鸟""金丝雀""鸽子""夜莺""鹦鹉"等。属性集或特征集是指一个概念的属性或特征,如"鸟"概念的特征为"是动物""有羽毛""有翅膀""会飞"等。这些特征称作语义特征。这样来看,语义记忆是由无数的这种信息集所构成的。然而,这些信息集或概念之间没有现成的联系。当要从语义记忆中提取信息来对句子作出判断时,如判断"金丝雀是鸟"这个句子的真伪,就可以分别搜索"金丝雀"和"鸟"的属性集,再对这两个属性集进行比较,根据这两个属性集的重叠程度作出决定。两个集的共同属性越多,重叠程度就越高。重叠程度高时,就可以作出肯定判断;反之则作出否定判断。由于"金丝雀"与"鸟"的属性集高度重叠,所以可迅速作出肯定判断。而对"金丝雀是动物"这个句子的判断,因"金丝雀"与"动物"的属性集也有相当高的重叠,故也可作出肯定判断,但其重叠程度低于"金丝雀"与"鸟"的属性集的重叠,因为"金丝雀"与"动物"的共同属性少于"金丝雀"与"鸟"的共同属性,所以作出判断就要慢些。可见,集理论模型也能说明范畴大小效应。但它是用两个概念的属性集的重叠程度来说明的,这与层次网络模型和激活扩散模型应用逻辑层次或连线都不同。

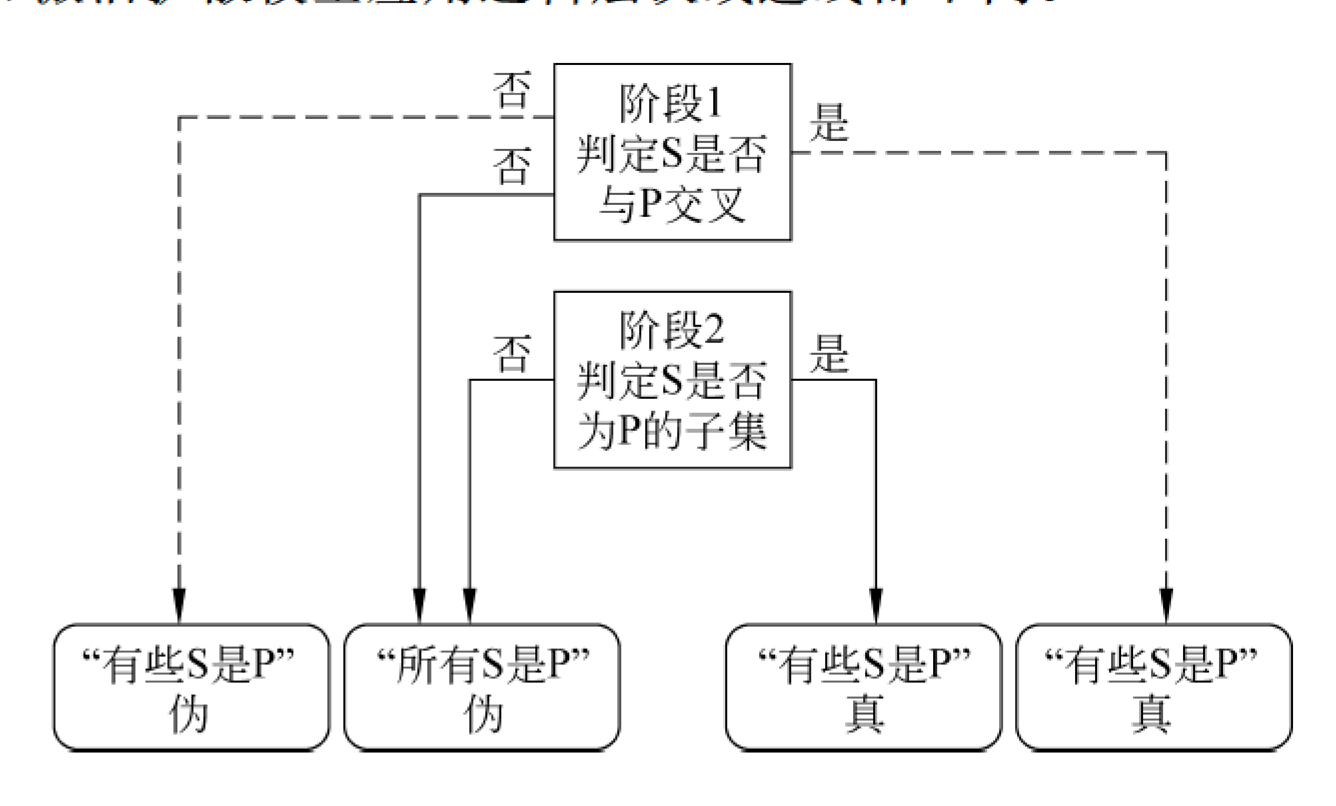



图 9.11 谓语交叉模型

### 4. 特征比较模型

特征比较模型是史密斯(E E Smith)等提出来的[723]。它同样被认为,概念在长时记忆中是由一集属性或特征来表征的。但它与集理论模型有一个很大的区别。集理论模型对一个概念的诸属性或特征没有按照其重要性加以区分,实际上,将它们看成对这个概念是同等重要的。特征比较模型则将一个概念的诸语义特征分成两类。一类为定义性特征,即定义一个概念所必须的特征。另一类为特异性特征,它们对定义一个概念并不必要,但也有一定的描述功能。图 9.12 列出"知更鸟"和"鸟"两个概念的特征及其比较。图 9.12 中所列特征是按其重要性自上而下排列的,位置越高也越重要。从图中可以看出,上级概念("鸟")具有的定义性特征比下级概念("知更鸟")要少,但下级概念的定义性特征必然要包含上级概念的全部定义性特征,此外还有自己的独特的特征。定义性特征和特异性特征可看作一个语义特征连续体的两端,语义特征的"定义性"即重要性的程度是连续变化的,可以任意选择一点将重要的特征与不太重要的特征分开来。特征比较模型强调定义性特征的作用。

特征比较模型认为,概念之间共同的语义特征特别是定义性特征越多,则其联系越紧密。

图 9.13 给出特征比较模型信息加工过程的两个阶段。第一阶段：提取命题的主语和谓语两个概念的特征，将两者的全部特征包括定义性特征和特异性特征加以总体比较，并确定两者的相似程度。如果两者高度相似则作出肯定反应，如果两者极不相似则作出否定反应。如果二者中等相似则进入第二阶段。第二阶段：撇开主语和谓语概念的特异性特征，只对两者的定义性特征进行比较、加工，如果两者匹配，则作出肯定反应，否则作出否定反应。两个加工阶段各有特点：第一阶段为总体比较，带有启发性质，常可发生错误；第二阶段加工为计算，较少发生错误。

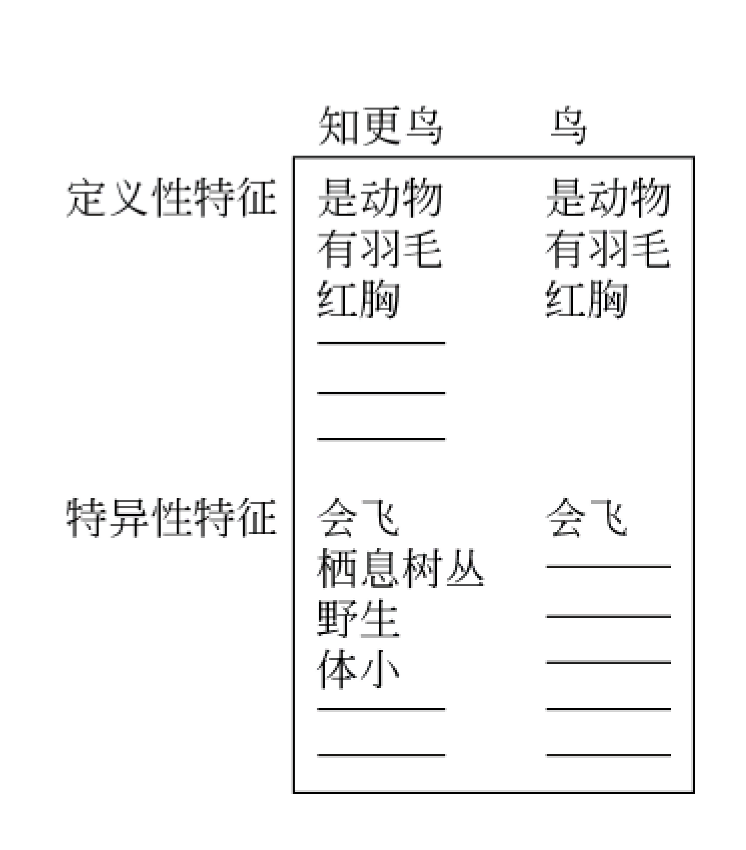



图 9.12　概念特征

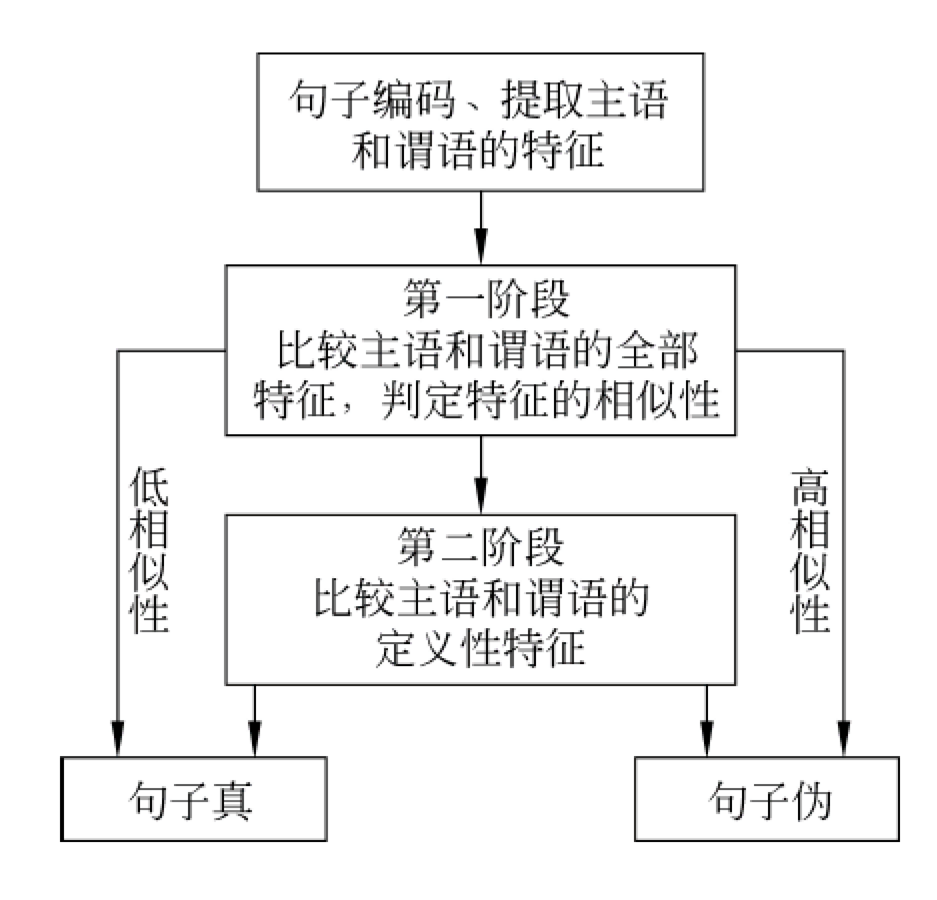



图 9.13　特征比较模型的加工阶段

特征比较模型可以解释典型性效应，以语义特征的相似性对各种实验结果进行解释，显得简洁有效。但它也有一个问题，即如何分清诸概念中定义特征与特异性特征的区别，而且也有一反例对该模型提出了质疑。语义记忆模型研究的困难在于语义记忆无法直接观察，需要通过操作过程才能推论出来。这种操作总表现在结构和过程两个方面。正因为对两者的关注点不同，才出现了众多的模型。

### 5. 人联想记忆模型

人联想记忆（Human Association Memory，HAM）模型的最大优点是既可以表征语义记忆，也可以表征情景记忆，既可以加工言语信息，也可以加工非言语信息，可以解释练习效应，也成功地实现过计算机模拟。但不能解释熟悉效应等现象，而其匹配过程是按阶段进行的思想，也受到质疑。认为语义记忆的基本表征单元是命题，而不是概念。

一个命题是由一小集联想构成的，每个联想将两个概念结合在一起。联想有四种类型：①上下文-事实联想，事实是指发生了什么事情，上下文是指何时何地发生了“何事”，它们结合而成上下文；②主语-谓语联想，主语说明事实的主体，谓语说明主体的特性；③关系-宾语联想，这是构成谓语的联想，关系是指主体的某种行动或与其他事物的联系，宾语则指行动的对象；④概念-实例联想，如家具-桌子。这几种联想的适当结合，就可以形成一个命题。用树形图可以很好地表明多种联想怎样结合而成一个命题。这种图可称作命题树。图 9.14 给出“教授在教室里问过了比尔”这个句子的命题树。从图中可以看到，命题树由结点（圆）和指针构成，结点代表命题、上下文、事实等概念。指针代表联想。图 9.14 的顶部结点 A 代表命题，称作命题结点，它是由事实和上下文之间的联想构成的。下面则是上下文

结点 B,它又包含地点结点 D 和时间结点 E(过去时,因为句中说教授问过了)。事实结点 C 也可分为两部分,即主语结点 F 和谓语结点 G;谓语结点再分为关系结点 H 和宾语结点 I。这个树形图的最底部则是终极结点,它们是长时记忆中的概念,即"教室""过去""教授""问""比尔"。因它们不可再分解,故称为终极结点。任何一个命题树必须连接它们才能获得一定意义。但是,概念不是按其本身的特性或概念的语义距离,而是按命题结构组织起来的,具有网络的性质,形成命题树。因此,长时记忆也就像一个庞大的命题树网络。从这里可以看到 HAM 模型的一个最大的特点和优点,这就是它既可以表征语义记忆,又可以表征情景记忆,能将两者结合起来。说来简单,只要情景记忆信息是以命题来表征的,在这种有层次的命题树网络中,就可以容纳个人经历的各种事件。这是前面谈过的几个语义记忆模型所做不到的。HAM 的命题树结构还可使一个命题嵌进另一个命题,把几个命题有机地结合在一起,构成一个复杂的命题。例如,可将"教授在教室里问过了比尔"和"这使考试按时结束"两个命题合成一个复杂的命题。这时前一个命题就是主语,后一个命题就是谓语,成为"教授在教室里问过了比尔而使考试按时结束"。这个复杂命题同样可以用命题树来表征。

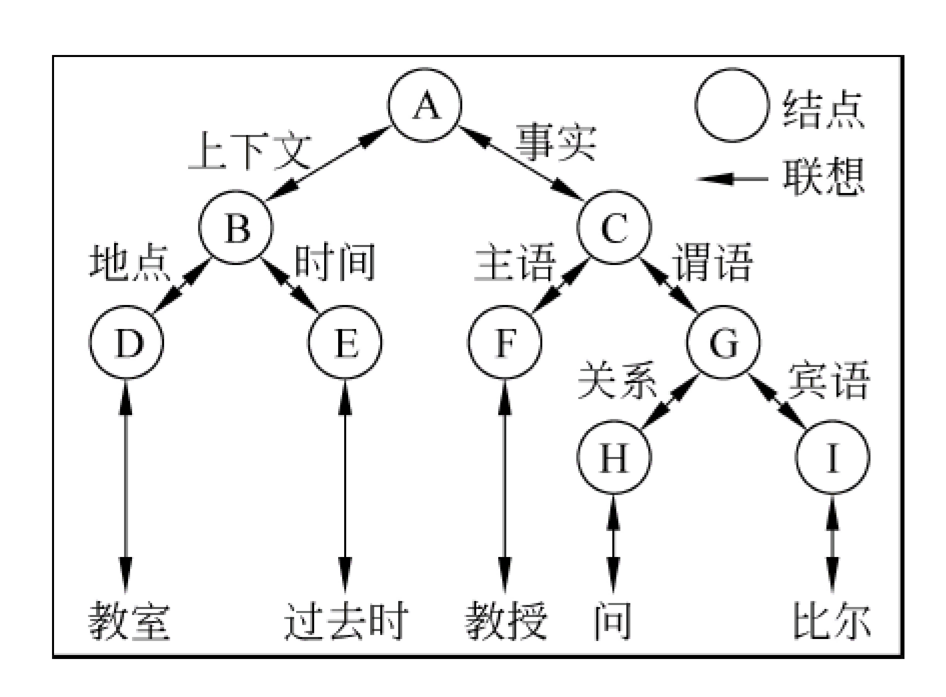



图 9.14　HAM 模型的命题树

HAM 模型认为,当需要从长时记忆中提取信息来回答一个问题或理解一个句子,其操作过程可分为四个阶段。

(1) 输入句子。

(2) 对输入的句子进行分析,构成一个命题树。

(3) 从长时记中的每个相应结点出发来搜索,以找到一个与输入的命题树相匹配的命题树。

(4) 搜索到的命题树与输入的命题树成功地匹配。

### 6. ELINOR 模型

ELINOR 模型取自该理论提出者 Lindsay、Norman 和 Rumelhart 三个人姓的头字母。该理论认为,长时记忆中存在三种信息类型:概念、事件和情景。概念指特定的思想,它由"是一个""有(是、会)""是一种(逆向)"三种关系来定义;事件是一个由行动、行动者和对象等构成的场景;事件是以行动为中心,围绕行动而展开的各种联系。由此,人的记忆是以事件为中心而组织起来的。概念是构成事件的成分。在 ELINOR 模型,所有的概念、事件和情景都用命题来表征。数个事件按一定时间关系结合而成情景,时间关系说明这些事件的先后顺序。例如,"母亲为孩子做早餐,孩子吃完早餐,背着书包上学"。在前面叙述的这 3

种类型的信息中，实际上起核心作用的是事件。可以说，ELINOR 模型的基本单元是事件。人的记忆是以事件为中心而组织起来的，概念是构成事件的成分。ELINOR 包含丰富的连线，除概念的 3 种关系以外，还有围绕行动而展开的各种关系。这使 ELINOR 可以表征复杂事物，对信息进行深入的加工，并且将语义记忆与情景记忆混合在一起。

ELINOR 模型是一个网络模型，其优点在于它可以容纳多种多样的联系，可表征各种信息，但它的加工过程尚不清楚，无法对其操作的结果作出预测。因而难以将它与其他模型作具体比较。

### 9.3.3　长时记忆的信息提取

长时记忆的信息提取有两种基本形式，即再认和回忆。

#### 1. 再认

再认（Recognition）是指人们对感知过、思考过或体验过的事物，当它再度呈现时，仍能认识的心理过程。再认与回忆没有本质的区别，但再认比回忆简单和容易。从个体心理发展来看，再认比回忆出现得早。孩子生后半年内，便可再认，而回忆的发展却要晚一些。日本学者清水曾用图画材料研究了小学生再认与回忆能力的发展。结果表明，幼儿园及小学低年级儿童的再认成绩明显优于回忆，而到五六年级时，两者的差别就逐渐趋向接近了。再认有感知和思维两种水平，并表现为压缩的和开展的两种形式。感知水平的再认往往以压缩的形式表现出来，它的发生是迅速而直接的。例如，对一首熟悉的歌队只要听见几个旋律就能立即确认无疑。思维水平的再认是以开展的形式进行的，它依赖于某些再认的线索，并包含了回忆、比较和推理等思维活动。再认有时会出现错误，对熟悉的事物不能再认或认错对象。发生错误的原因是多方面的。如接受的信息不准确；对相似的对象不能分化；有的错误则是由于情绪紧张或疾病等原因。

再认是否迅速和准确，要受到主客观方面许多因素的影响。重要的因素有以下几个方面。

(1) 再认依赖于材料的性质和数量。相似的材料，再认时容易发生混淆，如披与被，己与已等。材料的数量对再认也有影响。研究发现，在再认英文单词时，每增加一分词，再认时间就要增加 38%。

(2) 再认依赖于时间间隔。再认的效果随再认时间的间隔而变化。间隔越长效果越差。

(3) 再认依赖于思维活动的积极性。对于不熟悉的材料进行再认时，积极的思维活动可以帮助进行比较、推论、提高效果。例如，对一位多年不见的老朋友，可能记不起来了，这时根据现有线索，回忆过去的生活情景，能帮助对他的再认。

(4) 再认依赖于个体的期待。再认的速度和准确性不仅取决于对刺激信息的提取，而且依赖于主体的经验、定势和期待等。

(5) 再认依赖于人格特征。心理学家威特金（Witkinet）等人将人分为场依存性和场独立性。经过实验证实，具有场独立性的人不易受周围环境的影响，而具有场依存性的人易受周围环境的影响。这两种人，在识别镶嵌图形，即从复杂图形中识别简单图形时，有明显的差异。一般地说，场独立性的人比场依存性的人有较好的再认成绩。

### 2. 回忆

回忆是人们过去经历过的事物的形象或概念在人们的头脑中重新出现的过程。例如:考试时,人们根据考题回忆起学习过的知识;节日的情景,使人们想起远方的亲人。

在回忆过程中,人们所采取的策略,将直接影响回忆的进程和效果。

(1) 联想是回忆的基础。客观世界的各种事物不是孤立的,而是相互联系和相互制约的。人脑对客观事物的反映,在头脑中所保存的知识经验也不是孤立的和零散的,而是彼此有一定的联系的,这样人们在回忆某一事物时,也会连带地回忆起其他有关的事物。例如:想到"阴天"就会想到"下雨";想到一个朋友的名字,就会想到他的音容笑貌;等等。这种由一个事物想到另一个事物的心理活动称为联想。联想具有以下几个规律。

① 接近律:时间、空间相近的事物容易形成联想。例如,人们看到"颐和园"就会想到"昆明湖""万寿山""十七孔桥";背诵外文单词时由形会联想到它的音和义;由元旦会想到春节等。

② 相似律:形式相似和性质相似的事物容易形成联想。例如,人们提起春天,就会想到生机与繁荣;从苍松翠柏就会想到意志坚强;等等。

③ 对比律:事物间相反的特征也容易形成联想。例如,人们可能由白想到黑;由高想到矮;等等。

④ 因果律:事物间的因果关系也容易形成联想。例如,人们看到阴天就会想到下雨;看到冰雪就会想到寒冷;等等。

(2) 定势和兴趣直接影响回忆的方向和效果。定势对回忆有很大的影响,由于个人的心理准备状态不同,同一个刺激物可以使人回忆起不同的内容,产生不同的联想。另外,兴趣和情感状态也可以使人们对某一类事物的联想处于优势。

(3) 双重提取。寻找关键支点是回忆的重要策略。在回忆过程中,借助表象和词语的双重线索可以提高回忆的完整性和准确性。例如,问"家里有几扇窗户",首先在头脑中出现家中的窗户的形象,然后再提取窗户的数目,效果较好。在回忆中,寻找回忆材料的关键点,也有利于信息的提取。例如,回忆英文字母表,如果问字母表 B 后面的字母是什么?大部分人都能回忆起来,如果问 J 后面的字母是什么,回答就比较困难。在这种情况下,有的人从 A 开始通读字母表,知道 J 后面的字母是 K;而更多的人只从 G 或 H 开始,因为 G 在整个字母表上,形象比较突出,可能成为记忆材料的关键点。

(4) 暗示回忆和再认有助于信息的提取。在回忆比较复杂的和不熟悉的材料时,呈现与回忆内容有关的上下文线索,将有助于材料的迅速恢复。若暗示与回忆内容有关的事物,也能帮助回忆。

(5) 与干扰作斗争。在回忆过程中,经常会发生提取信息的困难,这可能是由于干扰所引起的。例如,考试时,有人明知考题的答案,但是由于当时情绪紧张,一时想不起来,这种明明知道而当时又回忆不起来的现象叫"舌尖现象",即话到嘴边又说不出来。克服这种现象的简便方法是当时停止回忆,经过一段时间后再进行回忆,要回忆的事物便可能油然而生。

# 9.4　工作记忆

1974 年巴德勒(A D Baddeley)等在模拟短时记忆障碍的实验基础上提出了工作记忆的概念[51]。传统的 Baddeley 模型认为工作记忆由语音回路、视觉空间画板两个附属系统和中枢执行系统组成[53]。语音回路负责以声音为基础的信息的储存与控制,包含语音储存和发音控制两个过程,能通过默读重新激活消退着的语音表征防止衰退,而且还可以将书面语言转换为语音代码。视觉空间画板主要负责储存和加工视觉空间信息,可能包含视觉和空间两个分系统。中枢执行系统是工作记忆的核心,负责各子系统之间以及它们与长时记忆的联系、注意资源的协调和策略的选择与计划等。大量行为研究和神经心理学上的许多证据表明了三个子成分的存在,有关工作记忆的结构和作用形式的认识也在不断地丰富和完善。

## 9.4.1　工作记忆模型

所有的工作记忆模型可以大致分成两大类,一类是欧洲传统的工作记忆模型,其突出代表就是巴德勒的多成分模型,强调把工作记忆模型分成多种具有独立资源的附属系统,突出通道特异性加工和储存。另一类是北美传统的工作记忆模型,以 ACT-R 模型为代表,强调工作记忆模型的整体性,突出一般性的资源分配和激活。前者的研究主要集中在工作记忆模型的储存成分,即语音回路和视空画板。如巴德勒明确指出,应该在探讨更复杂的加工问题之前,先把比较容易操作的短时储存问题研究清楚。而北美传统注重探讨工作记忆模型在复杂认知任务中的作用,如阅读和言语理解。因此北美传统所指的工作记忆模型类似于欧洲传统中一般性的中央执行系统。现在两种研究传统正越来越多地相互认同一些东西,并在各自的理论建构上产生相互影响。如情境缓冲区的提出与 Barnard 的"认知交互模型"中的命题表征系统很相似。因此,两大研究传统已表现出一定的整合和统一趋势。

巴德勒近年来有关工作记忆最大的发展是在传统模型的基础上,增加了一个新的子系统,即情境缓冲区[52]。巴德勒认为,传统的模型没有注意到不同类型的信息是怎样整合起来的,而且其整合结果是怎样保持的,因此不能解释随机的单词记忆任务中,被试者只能即时回忆出 5 个单词左右,但如果根据散文内容进行记忆,则能够回忆出 16 个左右的单词。情境缓冲区是一个能用多种维度代码储存信息的系统,为语音回路、视觉空间画板和长时记忆之间提供了一个暂时信息整合的平台,通过中央执行系统将不同来源的信息整合成完整连贯的情境。情境缓冲区与语音回路、视觉空间画板并列,受中央执行系统控制。虽然不同类型信息的整合本身由中央执行系统完成,但是情境缓冲区能保存其整合结果,并支持后续的整合操作。该系统独立于长时记忆,但却是长时情境学习中的一个必经阶段。情境缓冲区可用于解释系列回忆中的列表间位置干扰的问题、言语和视觉空间过程间的相互影响问题、记忆组块问题和统一的意识经验问题等。新增情境缓冲区之后的四成分模型如图 9.15 所示[52]。

罗夫特(Lovett)等的 ACT-R 模型则可用于解释大量个体差异方面的研究数据[505]。该模型把工作记忆资源看成一种注意激活,叫做源激活。源激活从当前的注意焦点扩散到与当前任务相关的记忆节点,并保存那些处于可获得状态的节点。ACT-R 是一个产生式系统,根据产生式规则的激活进行信息加工;强调加工活动对目标信息的依赖性,当前目标越

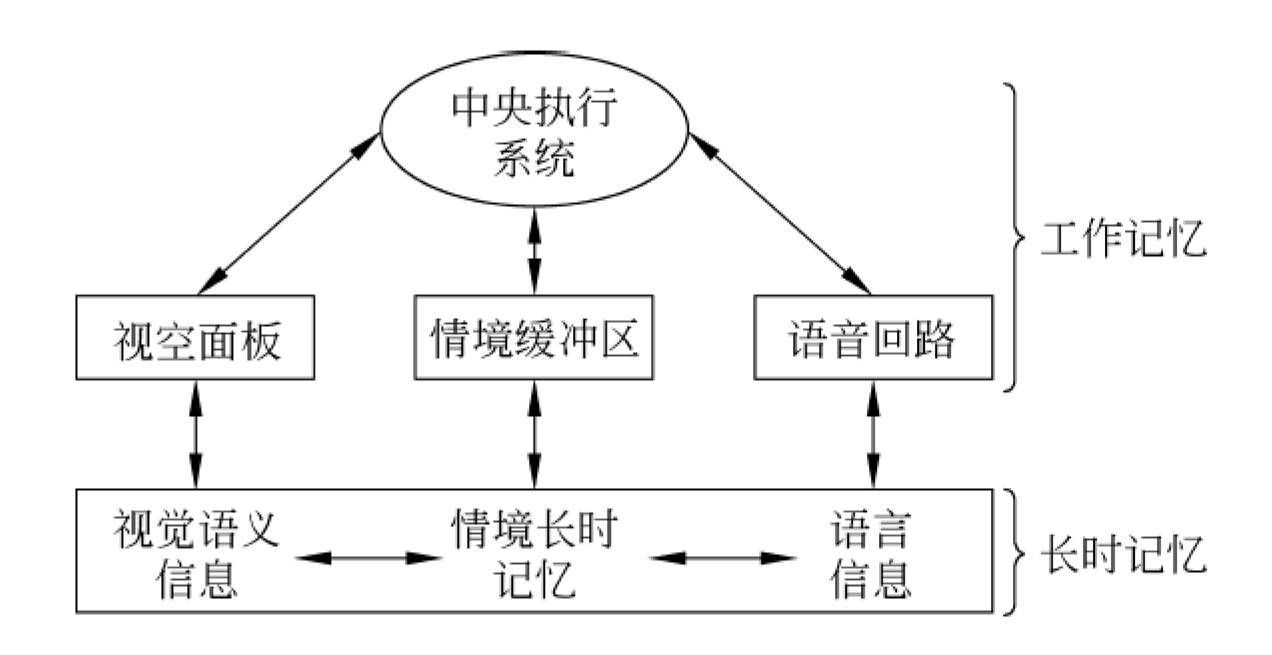



图 9.15 工作记忆的四成分模型

强烈,相关信息的激活水平就越高,信息加工就越迅速准确。该模型认为工作记忆容量的个人差异实际上反映了“源激活”总量的差异,用参数 $W$ 表示。而且这种源激活具有领域普遍性和单一性,语言和视觉空间信息的源激活基于相同的机制。该模型的明显缺陷在于只是用一个参数去说明复杂认知任务中的个体差异,因为工作记忆的个体差异还可能与加工速度、认知策略、已有知识技能有关,但 ACT-R 模型强调工作记忆的单一性,以详细阐明共同结构作为主要任务,能弥补强调工作记忆多样性的模型的不足。

## 9.4.2 工作记忆和推理

工作记忆与推理关系密切,工作记忆在推理中基本上有两个作用:一是保持信息;二是在工作记忆中形成初步的心理特征,中央执行系统的表征形式比两个子系统更为抽象一些。工作记忆是推理的核心,推理是工作记忆能力的总和。

根据工作记忆系统的概念,研究工作记忆各成分和推理之间关系一般采用“双重任务”实验范式。双重任务指的是同时进行两种任务,一个是推理任务,另一个是可以干扰工作记忆各成分的任务,称之为次级任务。干扰中央执行系统的活动是要求被试者随机产生字母或数字,或者利用声音吸引被试者的注意并做出相应行动,干扰语音环路采取的方法是要求被试者不断地发音,例如 the,the…或者按一定顺序数数,比如按 1,3,6,8 顺序数数等;对视觉空间初步加工系统干扰任务是持续的空间活动,比如被试者不看键盘,按一定顺序盲打。所有的次级任务都要保证一定的速率和正确率,并且与推理任务同时进行。双重任务的原理是两个任务同时竞争同一有限的资源。例如对语音环路的干扰使得推理任务和次级任务同时占用工作记忆子系统语音环路的有限资源,在这种条件下如果推理的正确率下降,时间延长,我们就可以确定语音环路参与了推理过程。有一系列的研究表明次级任务对工作记忆各成分的干扰是有效的。

吉尔霍利(Gilhooly)等研究了演绎推理和工作记忆的关系[265]。实验之一,发现呈现句子的方式会影响演绎推理的正确率,在视觉方式下的正确率比听觉方式时高,这是因为视觉方式对记忆的负荷低于听觉方式。实验之二,采用的是双重实验范式和视觉同时呈现句子的条件下,发现当有记忆负荷,即对中央执行系统进行干扰的情况下,演绎推理最容易受到影响和损害,次之是语音环路,视空加工系统最少参与。这表明在演绎推理中的表征是一种更为抽象的形式,符合推理的心理模型理论,导致了中央执行系统参与了推理活动。有可能语音环路也起了作用,因为与推理任务同时进行的语音活动减慢了,表明两种任务可能在竞争同一有限资源。在此实验中,吉尔霍利等人发现被试者在演绎推理中可能运用一系列的

策略，可以根据推理结果来推测被试使用的哪种策略。次级任务不同，被试使用的策略也可能不同，其对记忆的负荷也就不同；增加任务的负荷也会引起策略的变化，因为变化策略后对记忆的负荷也就降低了。1998 年吉尔霍利等人又用序列视觉呈现句子的方式采用双重任务实验范式研究工作记忆各成分和演绎推理的关系。序列呈现句子的方式比同时呈现句子的方式要求更多的存储空间。结果发现视觉空间系统和语音环路都参与了演绎推理，而且中央执行系统仍然在其中起着重要的作用。从以上结果可以得出结论，无论是序列呈现方式还是同时呈现方式，中央执行系统都参与了演绎推理；当记忆负荷增加时，有可能视觉空间系统和语音环路也参与推理过程。

### 9.4.3　工作记忆的神经机制

经过多年来，特别是近十年来脑科学的研究进展，已经发现思维过程涉及两类不同的工作记忆：一类用于存储言语材料（概念），采用言语类编码；另一类用于存储视觉或空间材料（表象），采用图形编码。进一步的研究表明，不仅概念和表象有各自不同的工作记忆，而且表象本身也有两种不同的工作记忆。事物的表象有两种：一种是表征事物的基本属性，用于对事物进行识别的表象，一般就称为“属性表象”或“客体表象”；另一种是用于反映事物空间结构关系（与视觉定位有关）的表象，一般称之为“空间表象”，或“关系表象”。空间表象不包含客体内容的信息，只包含确定客体空间位置或空间结构关系所需的特征信息。这样，我们就有三种不同的工作记忆。

（1）存储言语材料的工作记忆（简称言语工作记忆）：适用于时间逻辑思维。

（2）存储客体表象（属性表象）的工作记忆（简称客体工作记忆）：适用于以客体表象（属性表象）作为加工对象的空间结构思维，即通常所说的形象思维。

（3）存储空间表象（关系表象）的工作记忆（简称空间工作记忆）：适用于以空间表象（关系表象）作为加工对象的空间结构思维，即通常所说的直觉思维。

当代脑神经科学的研究成果已经证明，这三种工作记忆以及它们各自对应的思维加工机制，均可在大脑皮层中找到各自对应的区域（尽管有些工作记忆的定位目前还不很准确）。根据目前脑科学研究的新进展，布朗大学的布隆斯腾（S E Blumstein）指出[85]，言语功能并不是定位在一个狭小的区域上（按传统观念，言语功能只涉及左脑的布洛卡区和威尼科（Wernicke）区，而是广泛地分布于左脑外侧裂周围区域上，并向额叶前部和后部延伸，包括布洛卡区、紧邻脸运动皮层的下额叶和左侧中央前回（但不包括额极和枕极）。其中布洛卡区受损将影响言语表达功能，沃尼科区受损将影响言语理解功能。但是和言语理解与表达有关的加工机制并不仅仅限于这两个区。用于暂存言语材料的工作记忆一般都认为是在“左前额叶”，但具体是在左前额叶中的哪一部位，目前尚未精确定位。

与言语工作记忆相比，客体工作记忆与空间工作记忆的定位情况要准确得多。1993 年密歇根大学心理系的钟尼兹（J Jonides）等人运用当代研究脑科学的最先进测量技术之一正电子发射断层显像（Positron Emission Tomography，PET），对客体表象与空间表象的生成过程做了深入研究，得到了关于这两种表象生成机制与工作记忆定位的、富有价值的成果[351]。PET 是通过发射正电子的同位素作为标记物，将其引入脑内某一局部区域参与已知的生化代谢过程，然后用计算机断层扫描技术，将标记物参与代谢过程的代谢率以立体成像形式表达出来，因此具有定位准确、对大脑无损伤，适合于大量被试进行测试的优点。

## 9.5 遗忘理论

记忆是一种高级心理过程,受许多因素影响。旧联想主义者只是从结果推论原因,没有给予科学的论证。而艾宾浩斯(Hermann Ebbinghaus)则冲破冯特认为不能用实验方法研究记忆等高级心理过程的禁区,从严格控制原因来观察结果,对记忆过程进行定量分析,为此他专门创造了无意义音节和节省法。

旧联想主义者之间争论虽多,但对联想本身的机制结构从不进行分析。艾宾浩斯用字母拼成无意义音节作为实验材料,这就使联想的内容结构划一,排除了成年人用意义联想对实验的干扰。这是一项创造性工作,对记忆实验材料的数量化是一种很好的手段和工具。例如,他先把字母按一个元音和两个辅音拼成无意义的音节,构成 zog、xot、gij、nov 等共 2300 个音节,然后由几个音节合成一个音节组,由几个音节组合成一项实验的材料。由于这样的无意义音节只能依靠重复的诵读来记忆,这就创造出各种记忆实验的材料单位,使记忆效果一致,便于统计、比较和分析。例如,研究不同长度的音节组(7 个、12 个、16 个、32 个、64 个音节的音节组等)对识记、保持效果的影响以及学习次数(或过度学习)与记忆的关系等。

为了从数量上检测每次学习(记忆)的效果,艾宾浩斯又创造了节省法。它要求被试者把识记材料一遍一遍地诵读,直到第一次(或连续两次)能流畅无误地背诵出来为止,并记下诵读到能背诵所需要的重读次数和时间。然后过一定时间(通常是 24 小时)再学再背,看看需要重读次数和时间就能背诵,把第一次和第二次的次数和时间比较,看看节省了多少次数和时间,这就叫做节省法或重学法。节省法为记忆实验创造了一个数量化的统计标准。例如,艾宾浩斯的实验结果证明:7 个音节的音节组,只要诵读一次即能成诵,这就是后来被公认的记忆广度。12 个音节的音节组需要读 16.6 次才能成诵,16 个音节的音节组则要 30 次才能成诵。如果识记同一材料,诵读次数越多,记忆越巩固,以后(第二天)再学时节省下的诵读时间或次数就越多。

为了使学习和记忆尽量少受旧的和日常工作经验的影响,他应用了无意义音节作为学习、记忆的材料。他以自己做受试者,把识记材料学到恰能成诵,过了一定时间,再行重学,以重学时节约的诵读时间或次数,作为记忆的指标。他一般以 10～36 个音节,作为一个字表。在七八年间先后学了几千个字表。他的研究成果《记忆》发表于 1885 年。表 9.4 给出了他的实验结果的一例。利用表内材料可以划成一条曲线,一般称为遗忘曲线(图 9.16)。

**表 9.4 不同时间间隔后的记忆成绩**

| 时间间隔 | 重学节省诵读时间百分数 |
|---|---|
| 20 分钟 | 58.2 |
| 1 小时 | 44.2 |
| 8 小时 | 35.8 |
| 1 日 | 33.7 |
| 3 日 | 27.8 |
| 6 日 | 25.4 |
| 31 日 | 21.1 |

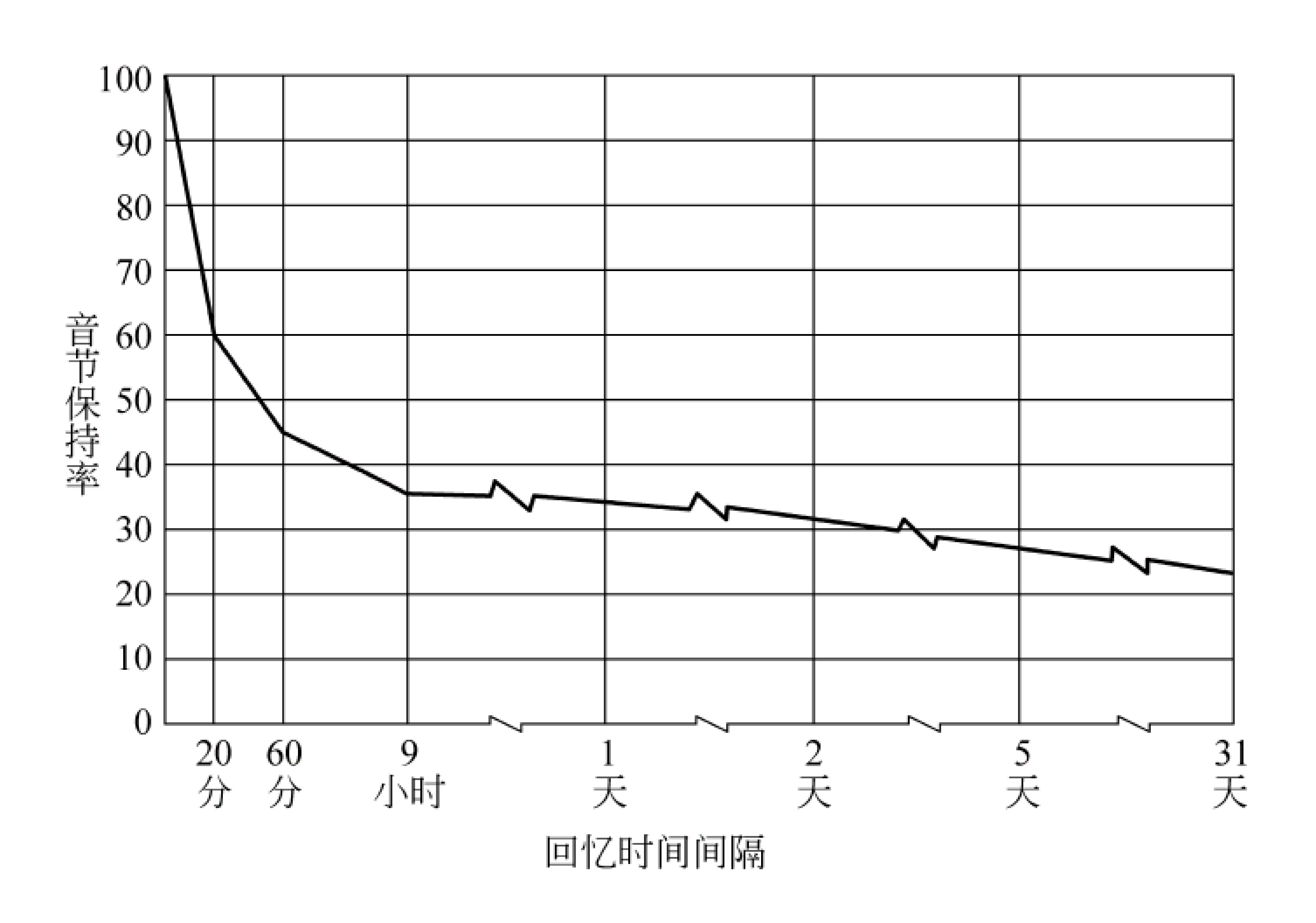



图 9.16 艾宾浩斯遗忘曲线

从艾宾浩斯的遗忘曲线中可以看到，一个明显的结果是，遗忘的过程是不均衡的：在第一个小时内，保存在长时记忆中的信息迅速减少，然后，遗忘的速度逐渐变慢。在艾宾浩斯的研究中，甚至在距初学 31 天以后，仍然存在着某种程度的节省，对所记的信息仍然有所保存。艾宾浩斯的开创性研究引发了两个重要的发现。一个是描述遗忘进程的遗忘曲线。心理学家后来用单词、句子甚至故事等各种材料代替无意义音节进行了研究，结果发现，不管要记的材料是什么，遗忘曲线的发展趋势都与艾宾浩斯的结果相同。艾宾浩斯的第二个重要发现是揭示了在长时记忆中的保存能够持续多长时间。通过研究发现，在长时记忆中信息可以保留数十年。因此，儿童时期学过的东西，即使多年没有使用，一旦有机会重新学习，都会较快地恢复到原有水平。如果不再使用，可能被认为是完全忘记，但事实上遗忘绝不是完全彻底的。

遗忘和保持是记忆矛盾的两个方面。记忆的内容不能保持或者提取时有困难就是遗忘，如识记过的事物，在一定条件下不能再认和回忆，或者再认和回记时发生错误。遗忘有各种情况：能再认不能回忆叫不完全遗忘；不能再认也不能回忆叫完全遗忘；一时不能再认或重现叫临时性遗忘；永久不能再认或回忆叫永久性遗忘。

对遗忘的原因，有各种不同的看法，归纳起来有下述四种。

### 1. 衰退说

衰退理论认为，遗忘是记忆痕迹得不到强化而逐渐减弱，以致最后消退的结果。这种说法易为人们所接受。因为一些物理的、化学的痕迹有随时间衰退甚至消失的现象。在感觉记忆和短时记忆的情况下，未经注意或重述的学习材料，可能由于痕迹衰退而遗忘。但衰退说很难用实验证实，因为在一段时间内保持量的下降，可能由于其他材料的干扰，而不是痕迹衰退的结果。有些实验已证明，即使在短时记忆的情况下，干扰也是造成遗忘的重要原因。

### 2. 干扰说

干扰理论认为,长时记忆中信息的遗忘主要是因为在学习和回忆时受到了其他刺激的干扰。一旦干扰被解除,记忆就可以恢复。干扰又可分前摄干扰与倒摄干扰两种。前摄干扰指已学过的旧信息对学习新信息的抑制作用,倒摄干扰指学习新信息对已有旧信息回忆的抑制作用。一系列研究表明,在长时记忆里,信息的遗忘尽管有自然消退的因素,但主要是由信息间的相互干扰造成的。一般说来,先后学习的两种材料越相近,干扰作用越大。对于不同内容的学习如何进行合理安排,以减少彼此干扰,在巩固学习效果方面是值得考虑的。

### 3. 压抑说

压抑理论认为,遗忘是由于情绪或动机的压抑作用引起的,如果这种压抑被解除了,记忆也就能恢复。这种现象首先是由弗洛伊德在临床实践中发现的。他在给精神病人施行催眠术时发现,许多人能回忆起早年生活中的许多事情,而这些事情平时是回忆不起来的。他认为,这些经验之所以不能回忆,是因为回忆它们时,会使人产生痛苦、不愉快和忧愁,于是便拒绝它们进入意识,将其储存在无意识中,也就是被无意识动机所压抑。只有当情绪联想减弱时,这种被遗忘的材料才能回忆起来。在日常生活中,由于情绪紧张而引起遗忘的情况,也是常有的。例如,考试时,由于情绪过分紧张,致使一些学过的内容,怎么也想不起来。压抑说考虑到个体的需要、欲望、动机、情绪等在记忆中的作用,这是前面两种理论所没有涉及到的。因此,尽管它没有实验材料的支持,也仍然是值得重视的一种理论。

### 4. 提取失败

有的研究者认为,储存在长时记忆中的信息是永远不会丢失的,我们之所以对一些事情想不起来,是因为我们在提取有关信息的时候没有找到适当的提取线索。例如,我们常常有这样的经验,明明知道对方的名字,但就是想不起来。提取失败的现象提示我们,从长时记忆中提取信息是一个复杂的过程,而不是一个简单的"全或无"的问题。如果没有关于某一件事的记忆,即使给我们很多的提取线索我们也想不出来。但同样,如果没有适当的提取线索,我们也无法想起曾经记住的信息。这就像在一个图书馆中找一本书,我们不知道它的书名、著者和检索编号,虽然它就放在书库中,我们也很难找到它。因此,在记忆一个词义的同时,尽量记住单词的其他线索,如词形、词音、词组和语境等,会帮助我们在造句时想起这个词。

在平常进行阅读时,信息的提取非常迅速,几乎是自动化过程。但有些时候,信息的提取需要借助于特殊的提取线索。提取线索使我们能够回忆起已经忘记的事情,或再认出储存在记忆中的东西。当回忆不起一件事情时,应该从多方面去寻找线索。一个线索对提取的有效性主要依赖于以下条件。

(1) 与编码信息联系的紧密程度。在长时记忆中,信息经常是以语义方式组织的,因此,与信息意义紧密联系的线索往往更有利于信息的提取。例如,触景生情,我们之所以浮想联翩是因为故地的一草一木都紧密地与往事联系在一起,它们激发了昔日的回忆。

(2) 情境和状态的依存性。一般来说,当努力回忆在某一环境下学习的内容时,人们往

往能够回忆出更多的东西。因为事实上我们在学习时，不仅将要记的东西予以编码，也会将许多发生在同时的环境特征编入了长时记忆。这些环境特征在以后的回忆中就成为有效的提取线索。环境上的相似性有助于或有碍于记忆的现象叫做情境依存性记忆。

同外部环境一样，学习时的内在心理状态也会被编入长时记忆，作为一种提取线索，叫做状态依存性记忆。例如，如果一个人在饮酒的情况下学习新的材料，而且测试也在饮酒的条件下进行，回忆结果一般会更好些。在心情好的情况下，人们往往回忆出更多美好的往事；而当人们心绪不佳时，往往更多记起的是倒霉事。

(3) 情绪的作用。个人情绪状态和学习内容的匹配也影响记忆。在一项研究中，让一组被试者阅读一个包含有各种令人高兴和令人悲伤事件的故事，然后在不同条件下让他们回忆。结果显示当人感到高兴时，回忆出来的更多的是故事中的快乐情境，而在悲哀时则相反。已有研究表明，心境一致性效应既存在于对信息的编码中，也包含在对信息的提取里。情绪对记忆的影响强度取决于情绪类型、强度和要记的信息内容。一般来说，积极情绪比消极情绪更有利于记忆，强烈的情绪体验能导致异常生动、详细、栩栩如生的持久性记忆。此外，当要记的材料与长时记忆中保持的信息没有多少联系时，情绪对记忆的作用最大。这可能是由于在这种情况下情绪是唯一可用的提取线索。

艾宾浩斯的研究是心理学史上第一次对记忆的实验研究，它是一项首创性的工作，为实验心理学打开了一个新局面，即用实验法研究所谓高级心理过程，如学习、记忆、思维等。在方法上力求对实验条件进行控制和对实验结果进行测量；激起了各国心理学家研究记忆的热潮，大大促进了记忆心理学的发展。艾宾浩斯虽然对记忆实验作出了历史性的贡献，但它也和任何新生事物一样，不可能是完美无缺的。其主要缺点是：艾宾浩斯对记忆过程的发展只作了定量分析，对记忆内容性质上的变化没有进行分析；他所用的无意义音节是人为的，脱离实际，有很大的局限性；他把记忆当作机械重复的结果，没有考虑到记忆是个复杂的主动过程。

## 9.6　内隐记忆

心理学对记忆的研究是沿着两条线路展开的：一条是传统的研究——注重对外显的、有意识的记忆研究；另一条便是对内隐的记忆的研究，是当前记忆研究的热点，代表了记忆研究的最新动向。

20 世纪 60 年代，瓦林顿(Warrington)和威斯克朗兹(Weiskrantz)在健忘症患者身上发现：他们没有意识到自己拥有对某方面的学习记忆，但在他们完成有关任务的操作上却表现出了记忆效果。这种现象被心理学家科菲(Cofer)称为启动效应(Priming Effect)。之后，对正常人进行大量研究发现，启动效应是普遍存在的，这是一种自动的、不需要有意识回忆的记忆现象。格雷夫(P Graf)和斯科特(D L Schacter)把这类记忆称为内隐记忆[281]，而把传统的、需经有意识回忆的记忆现象统称为外显记忆。

1987 年斯科特(D L Schacter)给出内隐记忆的早期表述是内隐记忆反映于那些先前经验易化了当前的任务操作，而该任务又不需要对先前经验的有意提取。2000 年麦卡德莫特(K B McDermott)认为内隐记忆是“在没有有意提取情况下的记忆显现”[485]。对内隐记忆的解释和建模工作一直以来都在进行，近年来在内隐记忆的模型方面，也有很大的发展和变

化。最为显著的是：以往内隐记忆的模型多为描述性的、定性的模型，其有效与否依赖于它们能否定性地预测实验结果；而最近的研究者开始发展可计算的、数量化的模型，模型与实验数据的拟合可以定量地表示。在这里介绍两个数量化的内隐记忆模型：REMI 和 ROUSE。

REMI 的全称是“记忆的有效提取(内隐)”(Retrieving Effectively from Memory, Imlplicit)。该模型假设：人们是以特征值的向量形式对学习项目进行表征的，可能被表征的特征有两个：内容信息和环境信息。在知觉辨识任务中，如果没有启动项目，则人们对目标项目和干扰项目的环境信息表征是一样的。因此人们的反应主要根据内容特征。依照贝叶斯推断进行优化决策。REMI 进一步假设，这种优化判断是基于备选项目间的“诊断特征”，也即项目间的不同之处进行的。判断过程首先分别计算每个项目能否符合所知觉内容的诊断特征，然后比较哪个项目的可匹配特征数量更多，胜出者即成为知觉辨识任务的反应。

然而启动项目的加入，使得启动项目除了在辨识任务中的内容特征表征之外(如果启动项目出现在辨识任务的备选项中)，更多了之前的环境特征信息表征。由于这些额外的环境特征同样可以匹配测验条件下的环境信息，因此启动项目的诊断特征匹配数量就会提高，使被试者的知觉辨识反应向着启动项目发生偏移。

REMI 可以预测长时启动如何随着知觉辨识任务中目标项和干扰项之间的相似性程度而变化。按照模型的预期，相似程度越小的各选项，即意味数量更多的诊断特征，进而直接导致两备选项目各自成功匹配的诊断特征数量之差将拥有更大的可能范围。如果将诊断特征匹配数之差为横坐标，以发生概率为纵坐标；那么高相似项目的辨识任务会产生一个高狭分布，而低相似任务产生扁平分布。虽然整个分布受到启动影响而水平位移的数量是相同的，但高狭分布因此而发生变化的正态曲线下、$x=0$ 标准以右的面积，无疑将大于扁平分布的对应面积变化。因此 REMI 预测相似项目条件受启动影响更大，而这一预测是和实际实验结果吻合的。

ROUSE 意指“利用不明来源证据的优化反应”(Responding Optimally with Unkown Sources of Evidence)。该模型对知觉辨识的短时启动机制解释，包括了三个参数：启动刺激以概率 $\alpha$ 激活目标项及干扰项中与之相关者的各个诊断特征(每个特征都为 0、1 二值，初始值为 0)；知觉辨识刺激呈现以概率 $\beta$ 激活目标项的各个特征；而神经系统和环境中的噪音以概率 $\gamma$ 激活目标项和干扰项的特征。

ROUSE 假设：①在启动刺激带来的激活与目标刺激闪现带来的激活间，会发生混淆，也即所谓的“不明来源证据”；②在这种混淆的前提下，人们为了做出优化反应，将对 $\alpha$ 的大小做出主观估计，并据此修正将启动刺激的影响从诊断特征的比较中去除。按此思路推理，如果 $\alpha$ 被高估，则被试者的判断反应将朝非启动项目发生偏移；而如果被低估，则被试倾向于选择启动项目。

那么在目标项和干扰项全部被启动或皆未被启动的情况下，结果将会如何？按照 ROUSE 的预测，由于目标项和干扰项在启动上的地位一致，被试者不会继续估计 $\alpha$ 值以及判断优化，而是直接做出判断。这样，在全部启动条件下，启动刺激同时加到目标和干扰项上的效应 $\alpha$ 就和 $\gamma$ 这样的噪音无异，而噪音增强的结果将导致目标正确辨识率的下降。

在 Huber 等(2001)对模型的实验检验中，消极启动组被试会被呈现两个启动词

500ms,随后紧接一个极短暂的目标词闪现,然后是两个备选词的迫选知觉辨识任务;而积极启动组被试的启动刺激将保持呈现直到被试对启动词做出"是否有生命"的判断,随后才是知觉辨识程序。实验结果和 ROUSE 的预测值十分接近:对启动词的更多加工导致被试对启动词效应的高估,因而消极条件下被试偏爱被启动项目,在积极条件下却变而喜好非启动项目;同时在两种启动条件下,也都观察到了此前预计的目标和干扰项同时启动情况下,判断正确率下降的结果。

从内隐记忆研究发展的大方向看,构建定量模型仍然是人们更准确地把握内隐记忆,甚至更广义上知觉、学习和记忆过程的重要途径。

## 9.7　动态记忆理论

记忆是如何组织的?记忆是怎样从过去经验中获得智能而自动改变和增长的?为了解释这些记忆过程的问题,是由香克在 1982 年提出了动态记忆理论[650]。当证实过去的经验失败就存储新的经验,即由经验中学习,用联想及从联想中归纳的方法改变结构以适应预测的失败。动态记忆依赖持续运行的神经冲动,或者依赖于可能由重复到达的冲动所强化的神经元内某些活跃的代谢变化和电位变化的保持。通过联系神经元的封闭环(或神经冲动)可能是这种动态记忆机制,每个记忆项目依赖于特定的神经元环或网的活动。事实上人们的记忆容量比神经元多,因此,不同的记忆可能共用部分通路。如果冲动实际上没有留下长时持续的痕迹,那么活动一旦停止,记忆便完全地,不可改变地消失掉。

为了说明动态记忆理论,可把专家存储关于书籍知识的方法与图书馆的目录系统作比较。在图书馆中,先根据图书的内容分类,然后按类别把图书卡装入不同的盒子,图书卡上包括每本书的书名、作者和主题。这种图书目录系统不是动态的。当某些类别过时了或要增加某些新类别时,必须修改原来的分类,这种修改很困难,而且需要外部干预。

专家的知识则没有这类问题。当他的兴趣或主题知识改变时,他可以很容易地改变他自己的内部分类系统。专家是自觉的生物,他能讲述自己所知道的事物,必要时他还可以改变他所知道的事物目录的存储结构。专家是一个动态记忆系统。

对图书馆来说,分类还不是唯一的问题。图书馆需要物理空间。不同的主题要占用各自的建筑物或楼层。有些主题可能在开始时被忽略而事后又要为它安排位置。另一些主题虽然开始时为它准备了很大的空间,但却从未放入任何材料。要想知道应该把某一本书或与该书有关的信息放在何处,必须事先对空间分配有一定的了解。即只有在原有经验的基础上才能作出决定。但当原有的安排被修改后,原来关于这方面的经验就不能用了。而人却能较容易地处理新信息。人类可以在自己的记忆中为新信息找到位置,当然也能找到原有的信息。

由此可知,动态记录是一个可变的、可扩充的系统,它可以根据新的经验而自动改变存储结构,它还可从大量的遗传经验中抽象出重要的一般原则及例外情况来进行学习。

香克提出了一种记忆组织包 MOP(Memory Organization Package)的记忆结构。MOP 既是记忆结构也是处理结构。MOP 可以存放新的输入,又可提供预测。根据以前出现过的结构相似的事件进行预测。

记忆组织包 MOP 是由景物(Scene)组成的序列。景物也是一种记忆结构,它把在同时

发生的且具有同一共享目标的各种动作(Action)组合在一起。景物提供了一个由一般动作组成的序列。在景物中还存储了各种特殊的记忆,并标出它们与景物中一般动作的区别。MOP 含有一组指向既定目标的景物。但一个 MOP 中总是存有一个主要的景物,该景物的目标是由该 MOP 所组织的事件的主体。

MOP 是景物的组织者,而景物又是记忆的组织者。记忆是存储在景物中的。MOP 可提供关于不同的景物是如何联系在一起的信息,还有关于它的各个景物的具体纲目,称作着色。一个景物可以由许多 MOP 共享,这种共享景物的能力提供了一种作出有用的概括的机构。当一个预测失败时,由于预测源是共享结构的一部分,故把失败及其预测都存储在这一结构中。当出现类似情况时,即使是在使用另一个 MOF,原先的失败由于联想机制仍会出现,因为这个失败是存储在共享的景物中的。预测失败可用来帮助建立新的 MOF;其方法就是改变原有 MOP 中景物的着色或顺序。

## 9.8 记忆-预测理论

霍金斯相信智能是大量群集的神经元涌现的行为,用基于记忆的世界模式产生连续不断的对未来事件的一系列预测。2004 年,他提出了记忆-预测理论,认为智能是以对世界模式的记忆和预测能力来衡量的,这些模式包括语言、数学、物体的物理特性以及社会环境。大脑从外界接收模式,将它们存储成记忆,然后结合它们以前的情况和正在发生的事情进行预测[306]。

大脑的记忆模式为预测创造了充分条件,可以说智能就是基于记忆的预测行为。大脑皮层的记忆具有如下属性。

(1) 存储的是序列模式。

(2) 以自联想方法回忆模式。

(3) 以恒定的形式存储模式。

(4) 按照层次结构存储模式。

### 9.8.1 恒定表征

图 9.17 显示了识别物体的前 4 个视皮质区域,分别用 V1、V2、V4、IT 表示。V1 表示条纹状视觉皮层区域,它对图像很少进行预处理,但包含着丰富的图像细节信息。V2 进行视觉映射,视觉图谱信息少于 V1。视觉输入用向上的箭头表示,始于视网膜,从图 9.17 中的底部开始传递到 V1 区。这个输入表示随时间变化的模式,由大约 100 万个神经轴突组成的视觉神经传输。

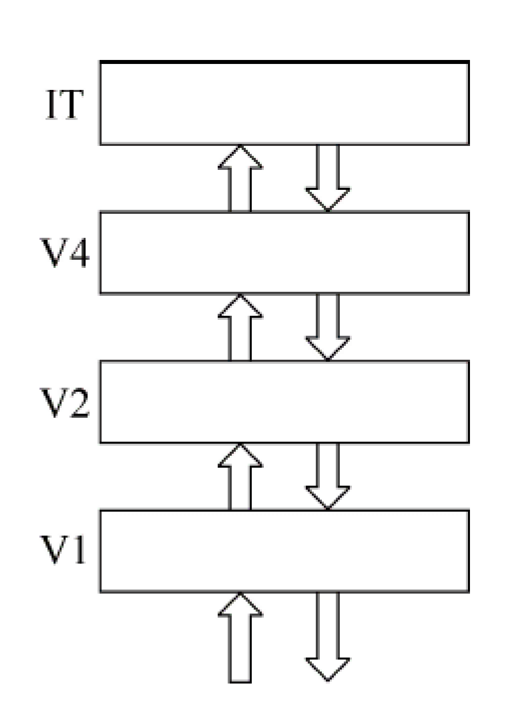


图 9.17 识别物体的前 4 个视皮质区域

在从视网膜到 IT 区的 4 个不同层次的区域中,细胞从快速变化、空间相关、能识别细微特征的细胞,逐渐变成了稳定激活、空间无关、能识别物体的细胞。例如,IT 细胞的“人脸细胞”,只要有人脸,就会被激活,不管出现的人脸是倾斜的、旋转的、还是部分被遮盖的,这是“人脸”的恒定表征。

当考虑预测时反馈连接很重要，大脑需要将输入信息送回到最初接收输入的区域。预测需要比较真正发生的事情和预期发生的事情。真正发生的事情的信息会自下而上流动，而预期发生的事情的信息会自上而下流动。

## 9.8.2　大脑皮层区的结构

大脑皮层的细胞密度和形状从上到下是有差异的，这种差异造成了分层。最顶部的第一层是 6 层中最独特的，包含的细胞很少，主要由一层平行于皮质表面的神经轴突组成。第 2、3 层比较类似，主要由很多紧挨在一起的金字塔形细胞组成。第 4 层由星形细胞组成。第 5 层既有一般的金字塔形细胞，还有一种特别大的金字塔形细胞。最下面的第 6 层也有几种独特的神经元细胞。

图 9.18 展示脑区的层次，同时一起协同工作的纵向细胞单元组成的垂直柱。每个垂直柱中的不同分层都通过上下延伸的轴突互相连接，并形成神经突触。在 V1 区的垂直柱有些对某方向的倾斜的线段(/)发生反应，而另一些会对朝另一个方向倾斜的线段(\)发生反应。每个垂直柱中的细胞都紧密互联，它们整体会对相同刺激产生反应。第 4 层的激活细胞会让在它之上的第 3、2 层的细胞激活。然后又会让它之下的第 5、6 层的细胞激活。信息在同一个垂直柱的细胞中上下传播。霍金斯认为，垂直柱是进行预测的基本单元。

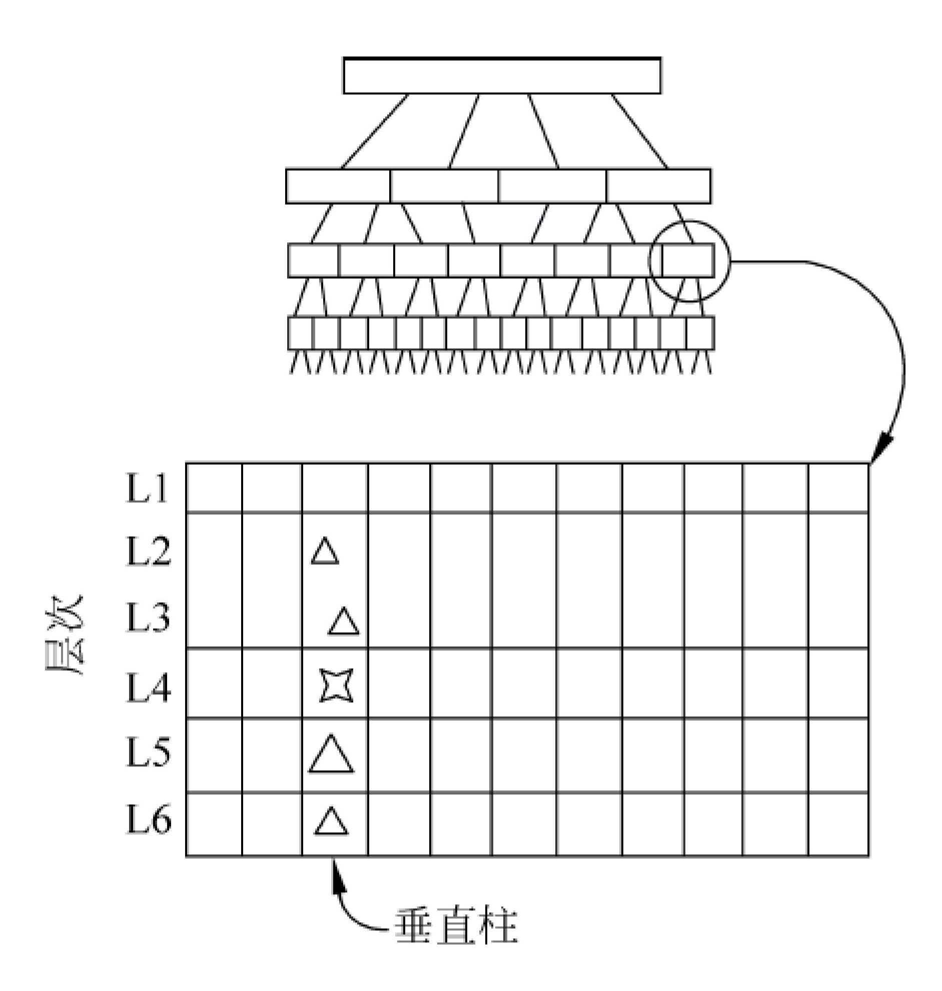



图 9.18　脑区的层次和垂直柱

运动皮层(M1)中的第 5 层细胞与肌肉以及脊髓中的运动分区存在着直接的联系。这些细胞高度协同地不断激活和抑制，让肌肉收缩，驱动运动。在大脑皮质的每个区域中都遍布第 5 层细胞，在各种类型的运动中都发挥作用。

第 5 层细胞的轴突进入丘脑区，连接到某类非特定的细胞上。这些非特定的细胞又会将轴突投射回大脑皮质不同区域的第 1 层中。这个回路正像自联想记忆中能够学会形成序列的延时反馈。第 1 层承载着大量信息，包括序列的名字以及在序列中的位置。利用第 1 层的这两种信息，一个皮层区域就能够学习和回忆模式序列了。

## 9.8.3　大脑皮层区如何工作

大脑皮层具有 3 种回路：沿皮层体系向上的模式会聚，沿皮层体系向下的模式发散，以及通过丘脑形成延时反馈，对大脑皮层区完成所需的功能极为重要。这些功能包括：

(1) 大脑皮层区如何将输入模式分类?

(2) 如何学习模式序列?

(3) 如何形成一个序列的恒定模式或者名字?

(4) 如何做出具体的预测?

大脑皮质的垂直柱中，来自较低区的输入信息激活了第 4 层细胞，导致该细胞兴奋。接

着第 4 层细胞激活了第 2、3 层细胞,然后是第 5 层,进而导致第 6 层细胞被激活。这样整个垂直柱就被低层区输入信息激活了。当其中一些突触随着第 2、3、5 层的激活而激活时,这些突触就会得到加强。如果这种情况发生足够多次,第 1 层的这些突触就会变得足够强,能够让第 2、3、5 层的细胞在第 4 层细胞没有激活的情况下也被激活。这样,第 2、3、5 层细胞就能根据第 1 层的模式预测应该何时激活。在这种学习前,垂直柱细胞只能被第 4 层细胞激活。而在学习之后,垂直柱细胞能够根据记忆获得部分的激活。当垂直柱通过第 1 层中的突触激活时,它就是在预测来自下方较低区的输入信息,这就是预测。

第 1 层接收的输入信息一部分来自相邻垂直柱和相邻区的第 5 层细胞,这些信息代表了刚刚发生的事件。另一部分来自第 6 层细胞,是稳定的序列名字。如图 9.19 展示,第 2、3 层细胞的轴突通常会在第 5 层形成突触,而从较低区到第 4 层的轴突也会在第 6 层形成突触。这两种突触在第 6 层的交集同时接收两种输入信息,就会被激活,根据恒定记忆作出具体预测。

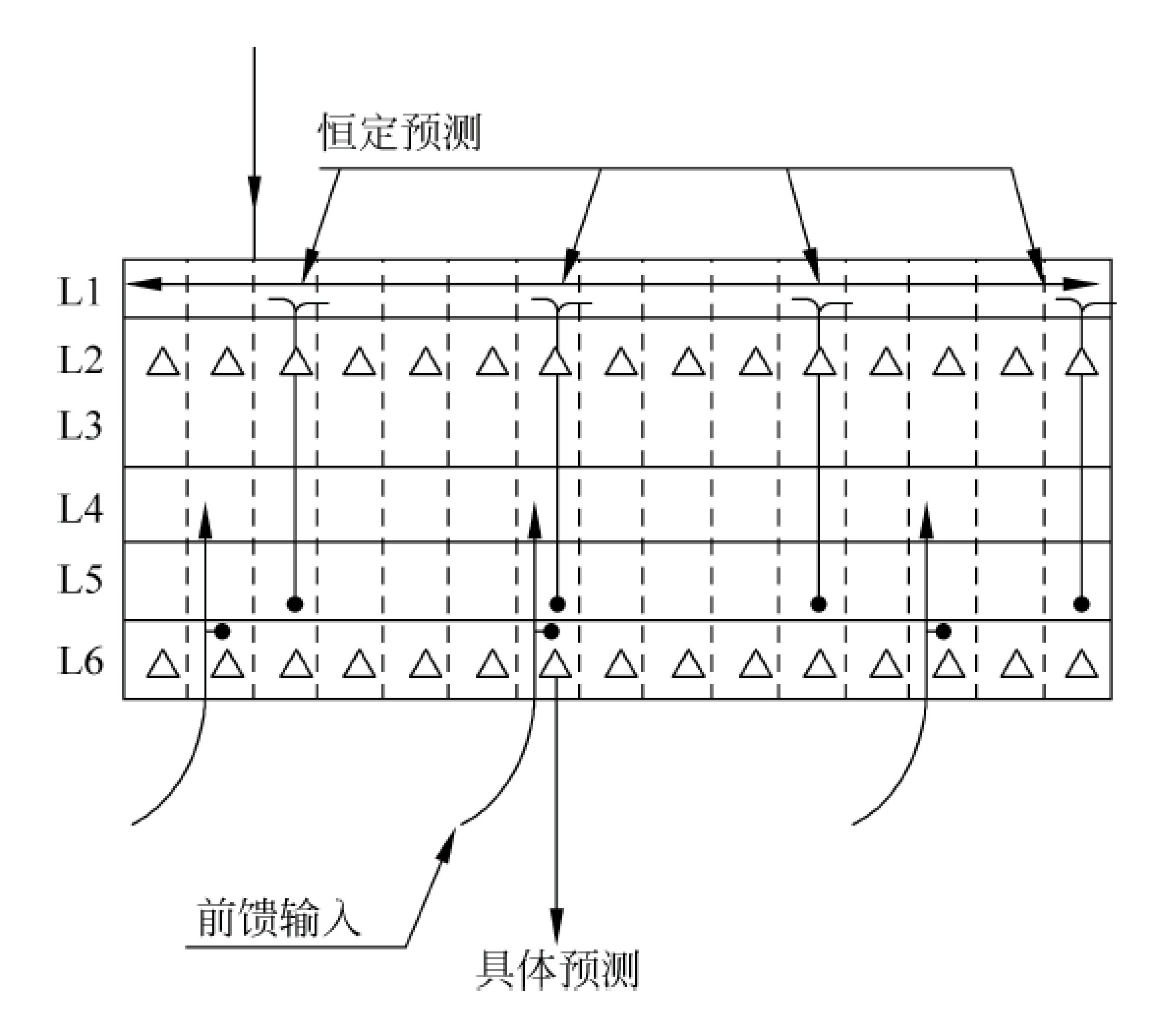



图 9.19　根据恒定记忆大脑皮层区作出具体预测

作为普遍规律,沿着大脑皮层向上流动的信息是通过细胞体附近的突触传递的。因此,向上流动的信息在传递过程中越来越确定。同样,作为普遍规律,沿着大脑皮层向下流动的反馈信息是通过细胞体远处的突触传递的。远距离的细树突上的突触能够在细胞激活中扮演积极且具有高度特异性的角色。通常情况下,反馈的轴突纤维要比前馈的多,反馈信息能够迅速准确地引起大脑皮层第 2 层中多组细胞激活。

# 9.9　互补学习记忆

## 9.9.1　海马体

一般认为,记忆的生理基础与新皮质和海马有关。端脑表面所覆盖的灰质称为大脑皮层。依据进化,大脑皮层分为古皮层(Archeocortex)、旧皮层(Paleocortex)和新皮层。古皮层与旧皮层与嗅觉有关,是三层的皮层,总称为嗅脑。人类新皮层高度发达,约占全部皮层

的 96%。新皮层发展成为六层，如图 9.20 所示。第一层是皮质内部神经元投射信息交汇的地方。底下 L2/3 和 L5 层的锥体细胞投射上来轴突和顶树突，在这里交汇，这里的神经元细胞很少，其中大部分都是抑制性的。L2/3 层有各种神经元，主要是小锥体细胞，构建皮层内的局部回路，这些锥体细胞主要连接是在皮质内部，但也有连到胼胝体的。L4 主要是颗粒性细胞，胞体较小而密集，负责接收丘脑传递的感觉信号。L5 主要负责传出信号，包含了最大的锥体细胞，将轴突投射到其他不同的脑区。L6 也是主要负责传出信号，但也接收丘脑传入的反馈信号。

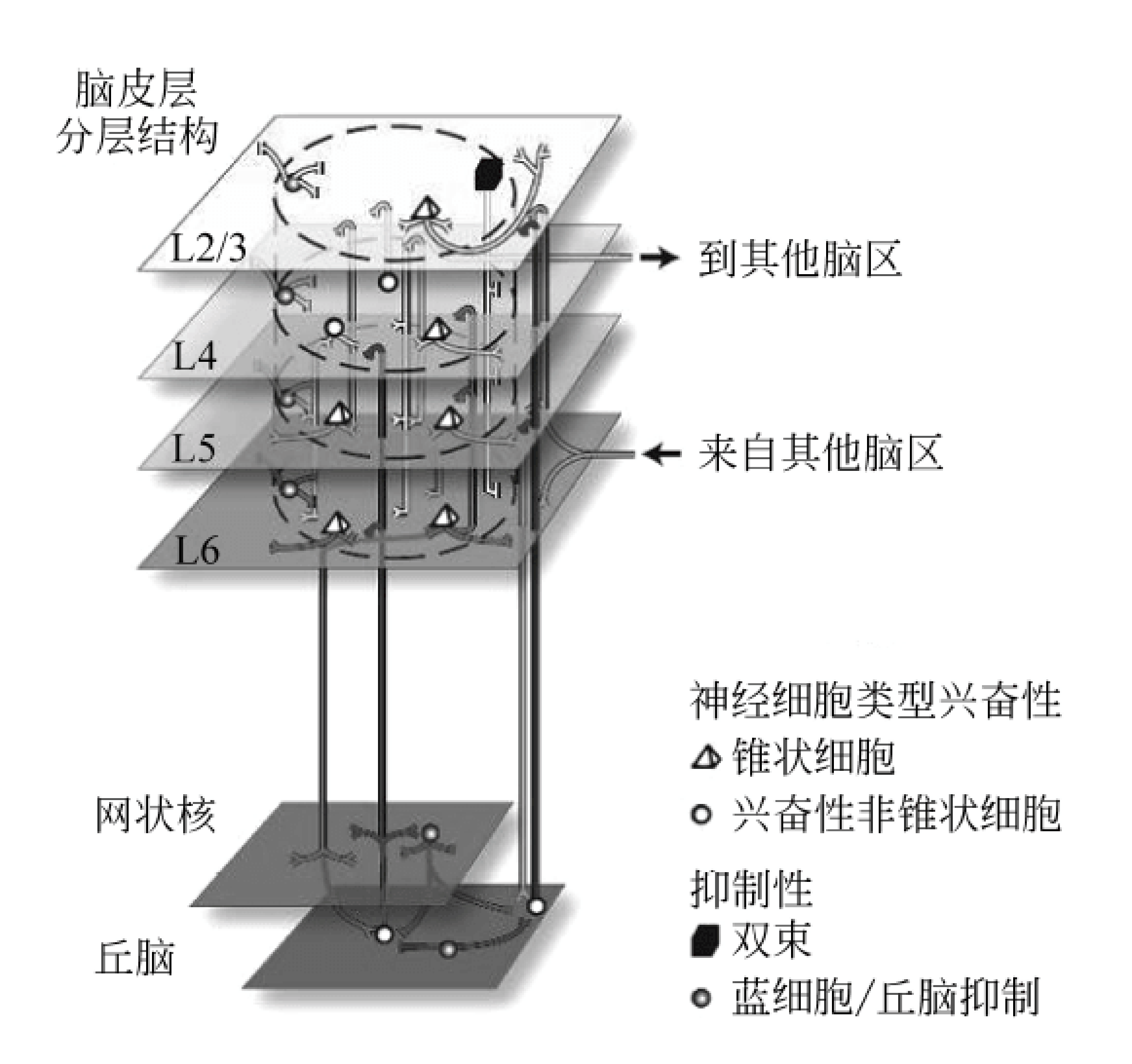



图 9.20　新皮层的分层结构

新皮层记忆结构化的知识，存储在新皮层神经元之间的连接中。当多层神经网络训练时，逐渐学会提取结构，通过调整连接权值，使网络输出的误差最小化，成为相对稳定的长时记忆。

海马体是大脑内部一个大的神经组织，它处于大脑半球内侧面皮层和脑干连接处。海马体由海马、齿状回和海马台组成。海马呈层形结构，没有攀缘纤维，而有许多侧枝。构成海马的细胞有两类，即锥体细胞和蓝细胞。在海马中，锥体细胞的细胞体组成层状并行的锥体细胞层，它的树突是沿海马沟的方向延伸。蓝细胞的排列非常有序。图 9.21 给出了海马体的构造。

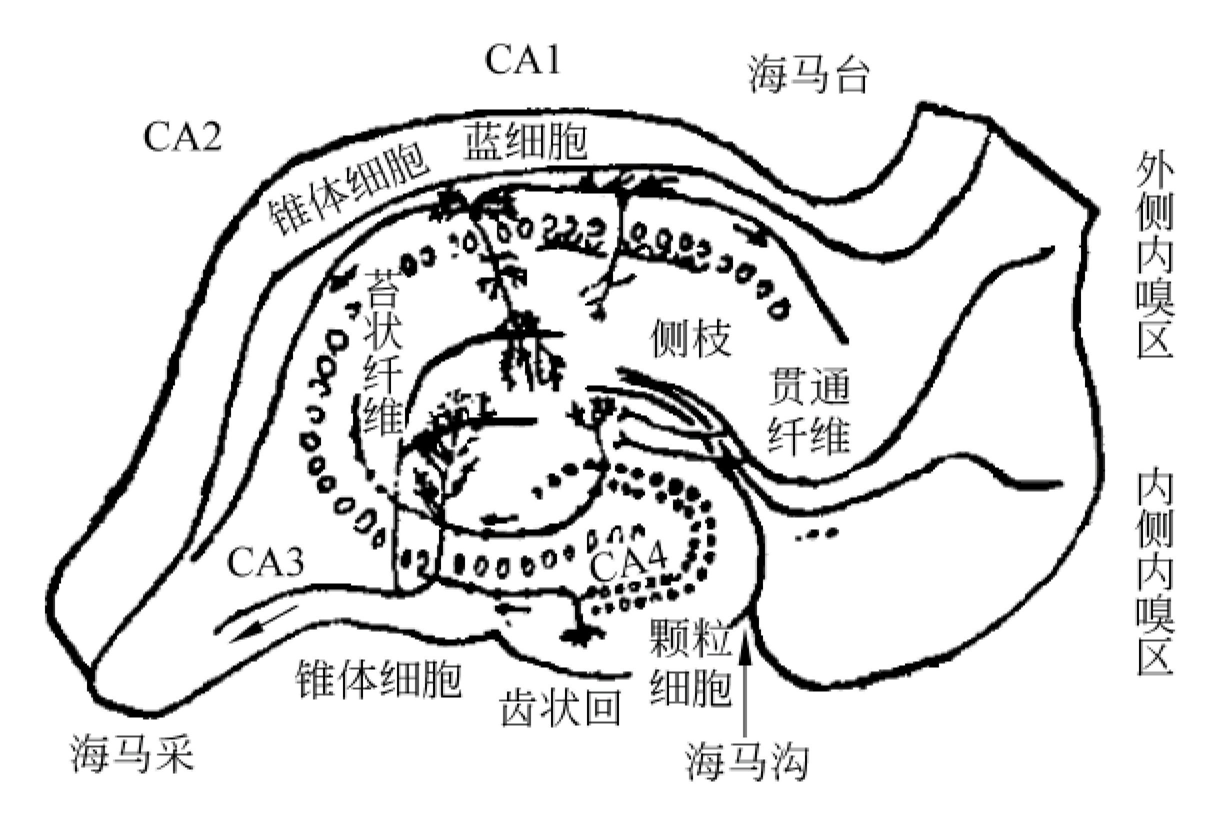



图 9.21　海马体的构造

海马区在存储信息的过程中扮演着至关重要的角色。短时记忆储存在海马体中。如果一个记忆片段，比如一个电话号码或者一个人在短时间内被重复提及的话海马体就会将其转存入大脑皮层，成为长时记忆。存入海马体的信息如果一段时间没有被使用，就会自行被“删除”，也就是被忘掉了。而存入大脑皮层的信息也并不是完全永久的，当你长时间不使用该信息的话大脑皮层也许就会把这个信息给“删除”掉了。有些人的海马体受伤后就会出现失去部分或全部记忆的状况。记忆在海马体和大脑皮层之间的传递过程要持续几周，并且

这种传递可能在睡眠期间仍然进行。

一些研究者运用 PET 技术来研究与陈述性记忆或外显记忆有关的大脑结构。当被试者完成陈述性记忆任务时右侧海马的脑血流量要比完成程序性记忆任务时的更高一些。这一发现支持海马结构在陈述性记忆中起到重要作用的观点。

## 9.9.2 互补学习系统

根据马尔早期的想法,斯坦福大学心理学教授麦克莱伦德(James L McClelland)于 1995 年提出了互补学习系统(Complementary Learning Systems,CLS)理论[478]。该理论认为人脑学习是两个互补学习系统的综合产物。一个是大脑新皮层学习系统,通过接受体验,慢慢地对知识与技能进行学习。另一个是海马体学习系统,记忆特定的体验,并让这些体验能够进行重放,从而与新皮层学习系统有效集成。2016 年,谷歌深度思维的库玛拉(Dharshan Kumaran)、哈萨比斯(Demis Hassabis)和斯坦福大学的麦克莱伦德在《认知科学趋势》刊物上发表文章[390],拓展互补学习系统理论。大脑新皮层学习系统是结构化知识表示,而海马体学习系统则迅速地对个体体验的细节进行学习。文章对海马体记忆重放的作用进行了拓展,指出记忆重放能够对体验统计资料进行目标依赖衡量。通过周期性展示海马体踪迹,支持部分泛化形式,新皮层对于符合已知结构知识的学习速度非常迅速。最后,文章指出了该理论与人工智能的智能体设计之间的相关性,突出了神经科学与机器学习之间的关系。

图 9.22 给出大脑半球的侧视图。其中虚线表示大脑内或深处的区域内侧表面,主要感觉和运动皮层显示为深黄色。内侧颞叶(Medial Temporal Lobe,MTL)包围虚线,海马以深灰色和周围的内侧颞叶皮质浅灰色(大小和位置是近似的)。绿色箭头表示整合的新皮层关联区域内和之间,以及在这些区域和模态特定区域之间的双向连接。蓝色箭头表示新皮质区域和内侧颞叶之间的双向连接。蓝色和绿色连接是互补学习系统理论中结构敏感的新皮层学习系统的一部分。内侧颞叶内的红色箭头表示海马内的连接,较浅红色的箭头表示海马之间的连接和周围的内侧颞叶皮质:这些连接表现出快速突触可塑性,这对将事件的元素快速结合成整合的海马表示非常重要。系统级合并涉及重播期间通过用蓝色箭头指示的途径扩散到新皮层关联区域的海马活性,从而支持在新皮层内连接(绿色箭头)内的学习。系统级合并是在记忆检索时完成,重新激活相关的新皮层表示集成,可以在无海马情况下完成。



图 9.22 大脑半球的侧视图

图 9.23 给出海马子区域、连通性和表示示意图,其中圆形或三角形表示神经元细胞体,红色线条表示高可塑性突触的投影,而灰色显示相对稳定的可塑性突触的投影。互补学习系统理论框架内的工作依赖于内侧颞叶分区的生理特性。在体验期间,来自新皮层的输入在内嗅皮层(Entorhinal Cortex,ERC)中产生激活的模式,可以被认为是压缩描述的贡献皮层区域中的模式。图 9.23 中内嗅皮层的说明性活动神经元以蓝色显示。内嗅皮层神经元产生投射到海马体的三个分区:齿状回(Dentate Gyrus,DG)、CA1 和 CA3。

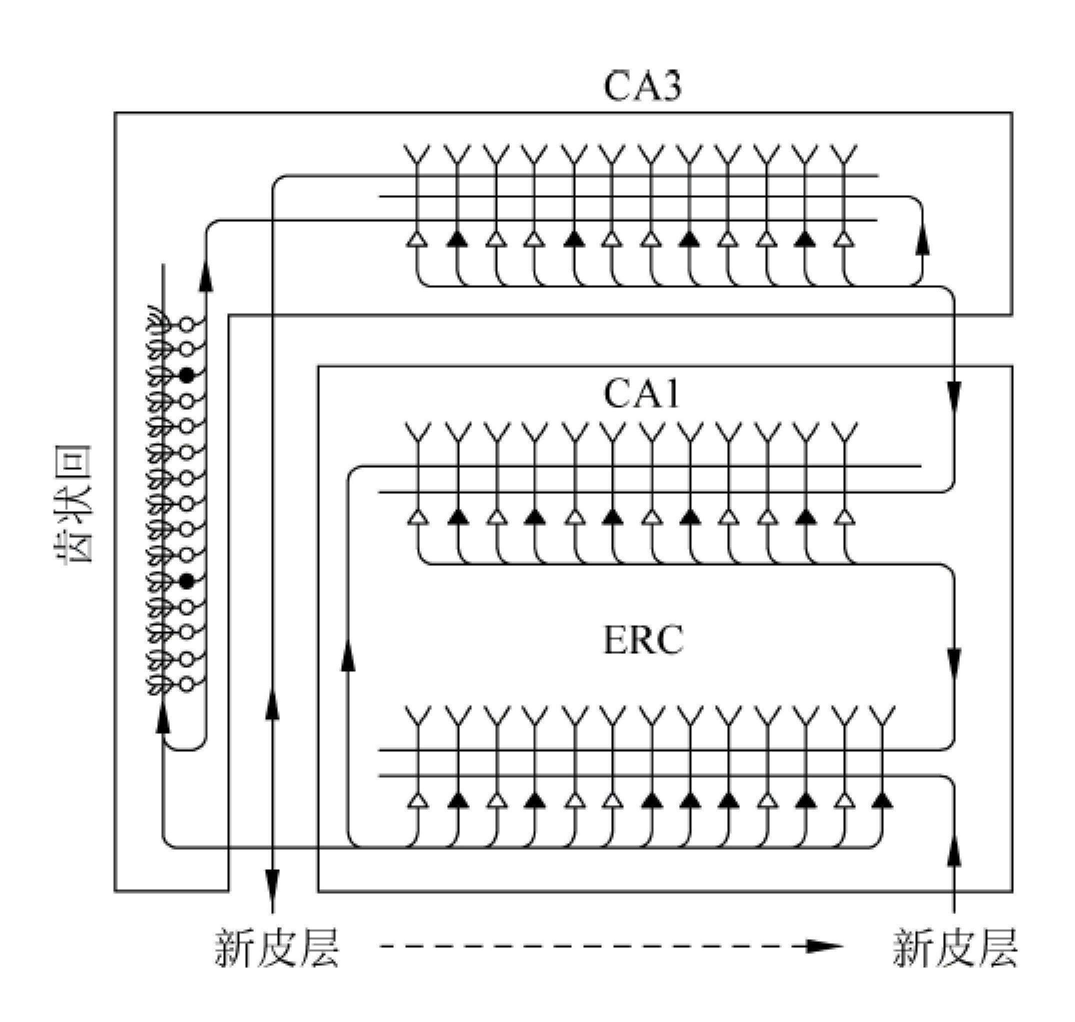



图 9.23　海马的子区域、连通性和表示的示意图

海马体学习系统实现模式选择和模式分离。新的内嗅皮层模式激活一组以前未提交的齿状回神经元，图 9.23 中显示红色，这些神经元可能是相对年轻的神经元，通过神经发生创建。这些神经元，反过来，通过大的“引爆突触”选择 CA3 中的神经元的随机子集。在从齿状回投影到 CA3 表示为红点，用作 CA3 中的记忆表示，确保新 CA3 模式与用于其他记忆的 CA3 模式尽可能不同，包括用于经验的模式，类似于新的经验。来自活动的 CA3 神经元反复连接到其他活动 CA3 神经元上，表示体验增强，使得相同神经元的子集稍后变为活动，其余的模式将被重新激活。从内嗅皮层到 CA3 的直接连接也得到加强，允许内嗅皮层输入在检索期间直接激活 CA3 中的模式，而不需要齿状回参与。

从内嗅皮层到 CA1 和背部的连接改变相对缓慢，允许 CA1 和内嗅皮层之间的模式相对稳定对应。在记忆编码期间，当 CA1 模式与相应的 CA3 模式被重新激活时，从活动的 CA3 神经元到活动的 CA1 神经元连接加强。从 CA1 到 ERC 的稳定连接则允许适当的模式有待重新激活，内嗅皮层和新皮层区域之间稳定的连接传播模式到大脑皮层。重要的是，CA1 和内嗅皮层之间，以及内嗅皮层和新皮层之间的双向投影，支持内嗅皮层和新皮层模式的可逆 CA1 表示的形成和解码，并允许重复计算。这些连接不应该快速改变给定的记忆中海马的扩展作用，否则在海马中存储的记忆在新皮层中难以恢复。

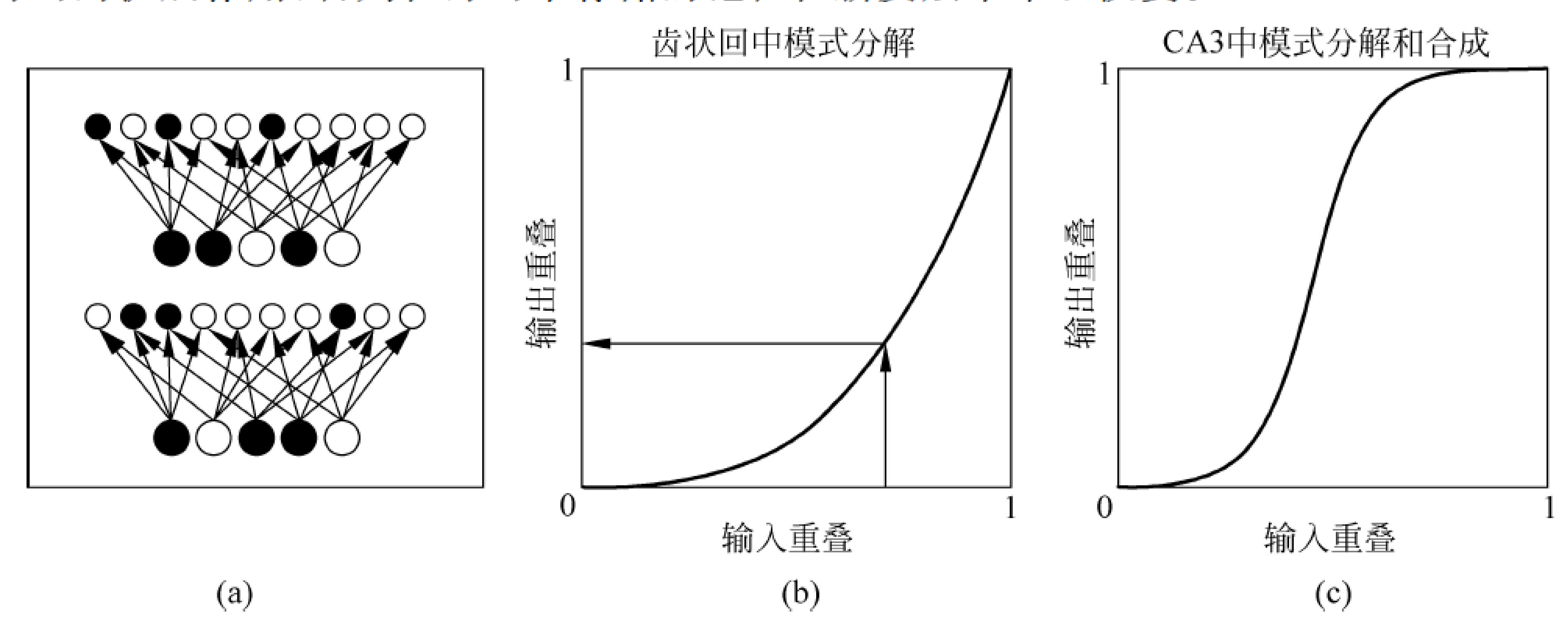



图 9.24　海马体连接的编码、模式分解和合成

图 9.24 解释海马的连接编码、模式分解和合成。图 9.24 (a)给出有 5 个输入端连接 10 个输出端,每个输出端连接 2 个不同模式的输入。每个连接单元检测相邻输入单元的活动。齿状回可以使用高阶的连接,放大这些影响。图 9.24 (b)说明模式分离函数的一般形式,显示在输入和输出之间的关系重叠。箭头表明图 9.24(a)中输入和输出的重叠。图 9.24(c)表示模式分解和合成的情况与 CA3 相关,在那里输入低的重叠减小少,而较高的输入重叠减小多。

模式分解和合成是根据影响神经活动模式之间的重叠或相似性的变换来定义的。模式分解使得类似模式通过连接编码更清晰,其中每个输出神经元仅响应于有效输入神经元的特定组合。图 9.24(a)和(b)显示了这种情况如何发生。模式分解是在齿状回中实现的,使用高阶连接减少重叠。

模式合成是一个过程,需要采集模式的片段和填补其余的功能。采集与熟悉模式相似的模式,使它更加类似于它。计算模拟已经表明 CA3 区域组合分离的模式特征和合成,使得中等和高度重叠,导致模式合成并保存在记忆中,而重叠较少的模式导致创建一个新的记忆,如图 9.24(c)所示。这种情况下,当环境输入在内嗅皮层中产生类似于先前模式的模式时,CA3 输出更接近其先前已在内嗅皮层模式的模式。然而,当环境在内嗅皮层上产生与以前记忆模式重叠低的输入时,齿状回在 CA3 中创建新的、统计独立的细胞群。新出现的证据建议模式合成(以及海马处理的其他特征)所需的重叠量可以发生在海马近端末梢和背腹侧轴,并可能由神经调节因子(如乙酰胆碱)形成。研究指出,CA3 和 CA1 分区之间的差异在于它们的神经活动模式对环境变化的响应。广义上讲,CA1 分区倾向于反映来自在内嗅皮质的输入重叠程度,而 CA3 更多显示反映模式分解或合成的不连续响应。

# 第 10 章<<
# 思维

思维是客观现实的反映过程，这个过程构成了人类认识的高级阶段。思维提供关于客观现实的本质的特性、联系和关系的知识，在认识过程中实现着“从现象到本质”的转化。与感觉和知觉，即与直接感性反映过程不同，思维是对现实的非直接的、经过复杂中介的反映。思维以感觉作为自己唯一的源泉，但是它超越了直接感性认识的界限，并使人能够得到关于它的感觉器官所不可能感知的现实的那些特性、过程、联系和关系的知识。

## 10.1 概述

思维是具有意识的人脑对于客观现实的本质属性、内部规律性的自觉的、间接的和概括的反映。思维的本质是具有意识的头脑对于客体的反映。“具有意识的头脑”的含义是有知识的头脑，又是具有自觉摄取知识的习性的头脑。“对于客体的反映”是反映客体的内在联系和本质属性，而不是表面现象的反映。

思维最显著的特性是概括性。思维之所以能揭示事物的本质和内在规律性的关系，主要来自抽象和概括的过程，即思维是概括的反映。所谓概括的反映是说所反映的不是个别事物或其个别特征，而是一类事物的共同的本质的特性。思维的概括性不只表现在它反映客观事物的本质特征上，也表现在它反映事物之间本质的联系和规律上。例如，地球围绕太阳旋转，通过感知，人只能反映地球和太阳在空间上的关系。而像地球围绕太阳旋转这类事物内部的联系，则是通过思维活动才能获得的，是人经过实践活动，通过概括、判断、推理，才能获得的反映，这就是著名的万有引力定律。万有引力公式

$$F = G\frac{m_1 m_2}{r^2} \tag{10.1}$$

定量地表述了宏观世界中两个物体间相互作用的图景，即相互作用的力与两个物体的质量乘积成正比，与它们之间的距离的平方成反比。

所谓间接的反映，就是说不是直接地、而是通过其他事物的媒介来反映客观事物。首先，思维凭借着知识经验，能对没有直接作用于感觉器官的事物及其属性或联系加以反映。例如中医专家通过望、闻、问、切四诊所获得的种种信息就可以确定病人的症状和体征，通过现象揭示事物的本质和内在规律性的关系。

正是由于思维的概括的、间接的性质，通过思维、人就可以认识那些没有直接作用于人的种种事物或事物的属性，也可以预见事物的发展变化进程。人不能直接感知光的运动速度，但通过实验可以间接地推算出光速 30 万千米每秒，人可以掌握直接感觉领域以外的东西，思考的东西的领域比感知的东西的领域要广阔。假设、想像，都是通过思维的间接性作为基础的。思维的这种间接性，使思维能够反作用于实践，指导实践，变成科学和理论，并揭

示事物发展的可能性。

思维之所以有间接性,关键在于知识与经验的作用。没有知识经验作为中介,思维的间接性就无法产生。思维的间接性是随着主体知识经验的丰富而发展起来。当然,思维的向接性问题,也反映了思维与记忆的相互关系,有了记忆,人才能积累知识,丰富经验,为思维提供材料。

在心理学史上,真正把思维当作心理学专门研究课题的,则是从冯特的学生丘尔佩(O Kulpe)开始。他和他的学生在符兹堡大学对思维心理学进行了大量的研究,形成了符兹堡学派,也叫做思维心理学派。

格式塔心理学着重研究了思维,并且开始研究儿童思维。他们提出课题在思维活动中的作用。他们认为思维是一个过程,是由问题情境中的紧张而产生的。在问题情境中能否产生紧张,也就是能否构成思维过程的课题,这在思维活动中起动力作用。顿悟是重要的学习理论之一。格式塔心理学认为,思维过程从紧张到解除紧张,是由问题情境的不断改组而最后取得解决的。用他们的术语说,就是"完形"的不断改组,直到领悟了问题内在的相互关系,就产生了"顿悟"。

顿悟学说是格式塔心理学借以对抗桑代克的尝试错误学说的。柯勒在对黑猩猩学习的实验中,发现一系列特点:①黑猩猩常常出现很长的停顿,它们表现出迟疑不决,并环顾四周的环境;②停顿表现为它们前后行动的转折点,停顿前的盲目行动,犹豫困惑,与停顿后的顺序前进,目的明确,造成强烈的对比;③停顿或转折后出现了一个不间断的动作序列,形成了一个连续完整体,正确地解决了问题,取得了目的物。于是柯勒等认为高等动物和人类的学习、思维,根本不是对个别刺激作个别反应,而是对整个情境作有组织的反应的过程;不是由于盲目的尝试,而是由于对情境有所顿悟而获得成功的。所谓顿悟,就是领会到自己的动作为什么和怎样进行,领会到自己的动作和情境,特别是和目的物的关系。对创造性思维也进行了一定的研究。

行为主义认为思维是无声的语言,是行为。华生认为,思想只是自己对自己说话。斯金纳认为,思想仅仅是一种行为,语词的或非语词的,隐蔽的或公开的。行为主义并不承认思维是脑的机能,而是全身肌肉,特别是喉头肌肉的内隐的活动。在对儿童思维或学习的研究方法上,行为主义反对内省,提倡实验。行为主义采用条件反射研究儿童的再现性思维的发展,说明思维、学习等是通过条件反射所习得的。

瑞士"日内瓦学派"的创始人皮亚杰是当代一位最著名的儿童心理学家或发生认识论专家。皮亚杰一生从事儿童思维活动的研究,发表了 30 多本著作,100 多篇论著。他把认知、智力、思维、心理视为同义语。他把生物学、数理逻辑、心理学、哲学等方面的研究综合起来,建立了自己的结构主义的儿童心理学或者发生认识论。1955 年,他集合各国著名的心理学家、逻辑学家、语言学家、控制论学者、数学家、物理学家等共同研究儿童认知的发生发展问题,在日内瓦建立了"发生认识论国际研究中心"。

维果斯基是苏联"维列鲁"心理学派的创始人。他的《思维和语言》一书,是苏联思维发展心理学的一本指导性著作;在这本书中,他指出了思维的生活制约性,客观现实对思维的决定作用,并提出思维是人的过去经验参与解决面临的新问题,是人脑借助于言语实现的分析综合活动;在这本著作中,他对儿童,特别是学龄前早期儿童的思维形成条件提出了一些见解。他指出,儿童的脑所具有的自然的思维发展的可能性,是在成年人的调节下与环境发

生相互作用的过程中实现的。儿童对实体世界的关系是以对教育他的人们的关系为中介的。主要利用言语实现与人们的交际，是儿童思维发展的特殊条件。维果斯基指出，学生理解的发展表现在概念形成过程的完善化中，真正的概念的形成似乎只是从少年时期开始的。他注意到思维与其他心理现象的关系，特别是思维与情绪、情感因素的联系。维列鲁学派对皮亚杰的发生认识论进行了研究。维果斯基肯定了皮亚杰在儿童的言语和思维的发展理论上的贡献，对皮亚杰关于儿童的自我中心言语的观点提出批评。维果斯基和他的学生列昂节夫(А Н Леонтъев)和鲁利亚(А РЛурия)合作，对儿童的自我中心言语的功能与作用问题进行了实验和临床研究，证明自我中心言语产生的原因和它的作用，指出自我中心言语是为解决困难服务的，是形式上的外部言语与功能上的内部言语的结合。这种自我中心言语通过社会言语——自我中心言语——内部言语的图式，逐渐从外部言语向内部言语过渡，它在儿童对周围环境中的定向与调节自己的活动中起着重要的作用。

瑞士“日内瓦学派”的创始人皮亚杰是当代一位最著名的儿童心理学家或发生认识论专家。皮亚杰一生从事儿童思维活动的研究，发表了 30 多本著作，100 多篇论著。他把认知、智力、思维、心理视为同义语。他把生物学、数理逻辑、心理学、哲学等方面的研究综合起来，建立了自己的结构主义的儿童心理学或者发生认识论。1955 年，他集合各国著名的心理学家、逻辑学家、语言学家、控制论学者、数学家、物理学家等共同研究儿童认知的发生发展问题，在日内瓦建立了“发生认识论国际研究中心”。

1995 年，美国加州大学心理学系的若宾(Nina Robin)等人发表了一篇题为“前额叶皮层的功能和关系复杂性”的论文[627]，从“前额叶皮层”是控制人类最高级思维形式的神经生理基础出发，试图探索出人类最高级思维模型与脑神经机制之间的联系。若宾等人认为，人类思维对于事物的本质属性和事物之间内在联系规律性所作的反映，实际上可看成是对事物之间存在的各种关系所作出的反映。根据数理逻辑中谓词逻辑的表述方式，事物本身所具有的本质属性也可看成是一种最简单的关系——一元关系；事物之间的相互联系则可看成是 $n$ 元关系。$n$ 是关系的维度，$n$ 愈大，关系的复杂程度愈高。换言之，$n$ 可作为描述关系复杂性的指标。在此基础上，若宾等人提出了一种用于确定关系复杂性的理论框架。然后，又根据当代脑神经科学所取得的成就，对前额叶皮层结构与机能的新认识，把对不同水平关系复杂性的处理和前额叶皮层中不同部位的控制机能联系起来，从而使我们对人类高级思维过程的认识，不仅建立在心理学的基础之上，而且深入到大脑内部的神经生理机制，因而有更为科学、更为坚实的基础。若宾等人在其论文中并未使用创造性思维这个术语，而是采用“最高级思维形式”“最独特思维形式”或“高水平认知”等概念。从该论文力图处理最高复杂程度的关系以及对“最高级”“最独特”的强调来看，作者所说的“最高级思维”其本意应当是指“创造性思维”。不过，就该论文中关于“最高级思维”的实际含义来看，若宾等人所提出的、用于处理关系复杂性的理论框架，实质上是一种建立在脑神经科学基础上的逻辑思维。尽管它还不是创造性思维模型，但是它对真正创造性思维模型的建立将具有一定的启迪意义。

若宾等人的高级思维模型建立在他们提出的“关系复杂性”的理论框架基础上。所谓“关系复杂性”是由关系中独立变化维数的数量 $n$ 来决定，因而根据 $n$ 的大小值即可给出不同关系的复杂性水平。

水平 1——一维函数关系，描述事物具有某种属性(若宾称之为“归因图式”)。

水平 2——二维函数关系,描述两种事物之间的二元关系(若宾称之为“关系图式”)。

水平 3——三维函数关系,描述三种事物之间的三元关系(若宾称之为“系统图式”)。

水平 4——$n$ 维函数关系($n>3$),描述 $n$ 种事物之间的 $n$ 元关系(若宾称为“多系统图式”)。

若宾等人认为,人类用来解决实际问题的各种知识不外乎两大类:明确的关系知识和内隐的关系知识。明确的关系知识以有意识的、可一步步进行逻辑推理的思维加工为基础;内隐的关系知识则以无意识(即潜意识)的、快速的直觉思维加工为基础。若宾等人通过脑神经解剖和电生理测量证实,前额叶皮层的主要功能就是获取和运用“明确的关系知识”(换句话说,前额叶皮层是实现逻辑推理的主要神经生理基础)。“关系复杂性理论”则是专门处理“明确的关系知识”的理论,所以下面只对这类知识进行讨论。

若宾指出,即使是婴儿也能根据客观的整体相似性而把一个苹果同另外一个苹果归为一类;但是如果要求按事物相同的颜色(只考虑颜色属性而忽略其他属性),把红苹果和红积木归为一类,那就需要在年龄稍大一些才有可能。若宾认为,这是因为前者是使用“内隐的关系知识”的潜意识加工,后者则要求儿童应能把角色与角色的填充符区分开才能做到。如上所述,区分角色就是按某种属性对某一类事物进行概括的过程,这实际上是一种逻辑思维(尽管是最初步的逻辑思维)过程。婴儿出生不久可以具有整体、直觉思维,但是逻辑思维则要到少年时代乃至 16~17 岁才能最后形成。正是在这个意义上,若宾等人把逻辑思维称之为“最高级的思维形式”。

所谓关系复杂性理论框架,就是建立在谓词逻辑基础上、专门用于表征“明确的关系知识”的一套知识表征系统,利用该系统可以方便地确定当前所处理知识的复杂性水平(由最简单到最复杂分成 1、2、3、4 四个等级)。

在上述“关系复杂性”理论框架的基础上,若宾等人根据当代神经解剖学和脑细胞电生理测量的证据,对前额叶皮层的结构与功能作了较深入的研究,指出前额叶皮层主要包括主沟及其周围的背侧部、弓沟及其周围部位及眶额部等三个组成部分,每一部分所具有的功能都是处理事物之间的复杂关系,即实现逻辑思维所必不可少的。

(1) 主沟及其周围的背侧部:负责控制注意、工作记忆,制定计划,并对刺激-反应这类偶然性事件的学习有一定影响。换言之,若大脑的这一部位受损伤,这几种与逻辑思维密切相关的心理操作将无法执行。

(2) 弓沟及其周围部位:这一部位对于刺激-反应这类条件性偶然事件的学习起决定作用,尤其是对突发事件的反应及处理更是至关重要。

(3) 眶额部:负责选择性反应和情绪控制。这一部位受损,将影响选择性作业(从背景中选出目的物和抵制干扰的能力),使情绪波动、情感障碍乃至人格变异。

对于具有时间顺序性和目标指向性的行为控制,则需上述三个部分协同工作,才有保证。若宾等人还对思维过程中不同信息加工情况的关系复杂性水平作了具体分析。例如:如果必须在一定时间内顺序地对多种信息进行整合,就势必会增加作业的关系复杂性:首先,顺序性即时间上的分离有碍于将不同信息组成更大的组块,于是就需要将原来的信息分为若干个独立而相关的单元来进行加工,这就会出现二元关系(关系图式)、三元关系(系统图式)甚至 $n$ 元关系(多系统图式)。与此同时,在思维加工过程中,还要为每一维度的信息提供暂时的工作记忆,以便等待该 $n$ 元关系中最后一个维度信息的到来(这时才能对该 $n$ 元

关系作出处理）。显然，这将会大大增加前额叶皮层关于工作记忆和注意分配的负担。此外，若宾等人运用关系复杂性理论还对“顺序回忆作业”和“非顺序回忆作业”的关系复杂性水平作了定量的对比分析，其结论是前者远远高于后者。

总之，若宾等人建立在关系复杂性理论框架基础上的高级思维模型，由于有神经解剖学和电生理测量证据的有力支持，使人类对整个逻辑思维的心理操作过程有了更为深刻的认识。这是到目前为止，在基于脑神经科学的逻辑思维模型中给人印象较深的一个模型。

在国内，20 世纪 80 年代初，钱学森倡导开展思维科学的研究。“思维科学”（Noetic Science）这一概念在中国近代最早是南叶青于 1931 年在一篇题为《科学与哲学》的文章中提出来的。他把自然、社会和思维三种现象放在同一层面上进行了严格的界定，然后指出，自然、科学和思维的根本区别就在于“自然现象是不经过人的行为就已经存在的，社会现象是要经过人的行为才能够存在的。思维现象是未经过人的行为，因而未外化成事实的观念作用和观念形态”。

1984 年，钱学森提出思维科学的研究[908]，研究思维活动规律和形式，并把思维科学划分为思维科学的基础、思维科学的技术科学和思维科学的工程技术三个层次。思维科学的基础科学研究思维活动的基本形式——逻辑思维、形象思维和灵感思维，并通过对这些基本思维活动形式的研究，揭示思维的普遍规律和具体规律。因此，思维科学的基础科学可有若干分支，如逻辑思维学、形象思维学等。个体思维的累积和集合，构成社会群体的集体思维。研究社会群体集体思维的是社会思维学。

## 10.2　思维的形态

人类思维的形态主要有感知思维、形象（直感）思维、抽象（逻辑）思维和灵感（顿悟）思维。感知思维是一种初级的思维形态。在人们开始认识世界时，只是把感性材料组织起来，使之构成有条理的知识，所能认识到的仅是现象。在此基础上形成的思维形态即是感知思维。人们在实践过程中，通过眼、耳、鼻、舌、身等感官直接接触客观外界而获得的各种事物的表面现象的初步认识，它的来源和内容都是客观的、丰富的。

形象思维主要是用典型化的方法进行概括，并用形象材料来思维，是一切高等生物所共有的。形象思维是与神经机制的连接论相适应的。模式识别、图像处理、视觉信息加工都属于这个范畴。

抽象思维是一种基于抽象概念的思维形式，通过符号信息处理进行思维。只有语言的出现，抽象思维才成为可能，语言和思维互相促进，互相推动。可以认为物理符号系统是抽象思维的基础。

对灵感思维至今研究甚少。有人认为，灵感思维是形象思维扩大到潜意识，人脑有一部分对信息进行加工，但是人并没有意识到。也有人认为，灵感思维是顿悟。灵感思维在创造性思维中起重要作用，有待进行深入研究。

人的思维过程中，注意发挥重要作用。注意使思维活动有一定的方向和集中，保证人能够及时地反映客观事物及其变化，使人能够更好地适应周围环境。注意限制了可以同时进行思考的数目。因此在有意识的活动中，大脑更多地表现为串行的。而看和听是并行的。

根据上述讨论,作者提出人类思维的层次模型(参见图 10.1)[915]。图中感知思维是极简单的思维形态,形成初级的思维。形象思维以神经网络的连接论为理论基础,可以高度并行处理。抽象思维以物理符号系统为理论基础,用语言表述抽象的概念。由于注意的作用,其处理基本上是串行的。

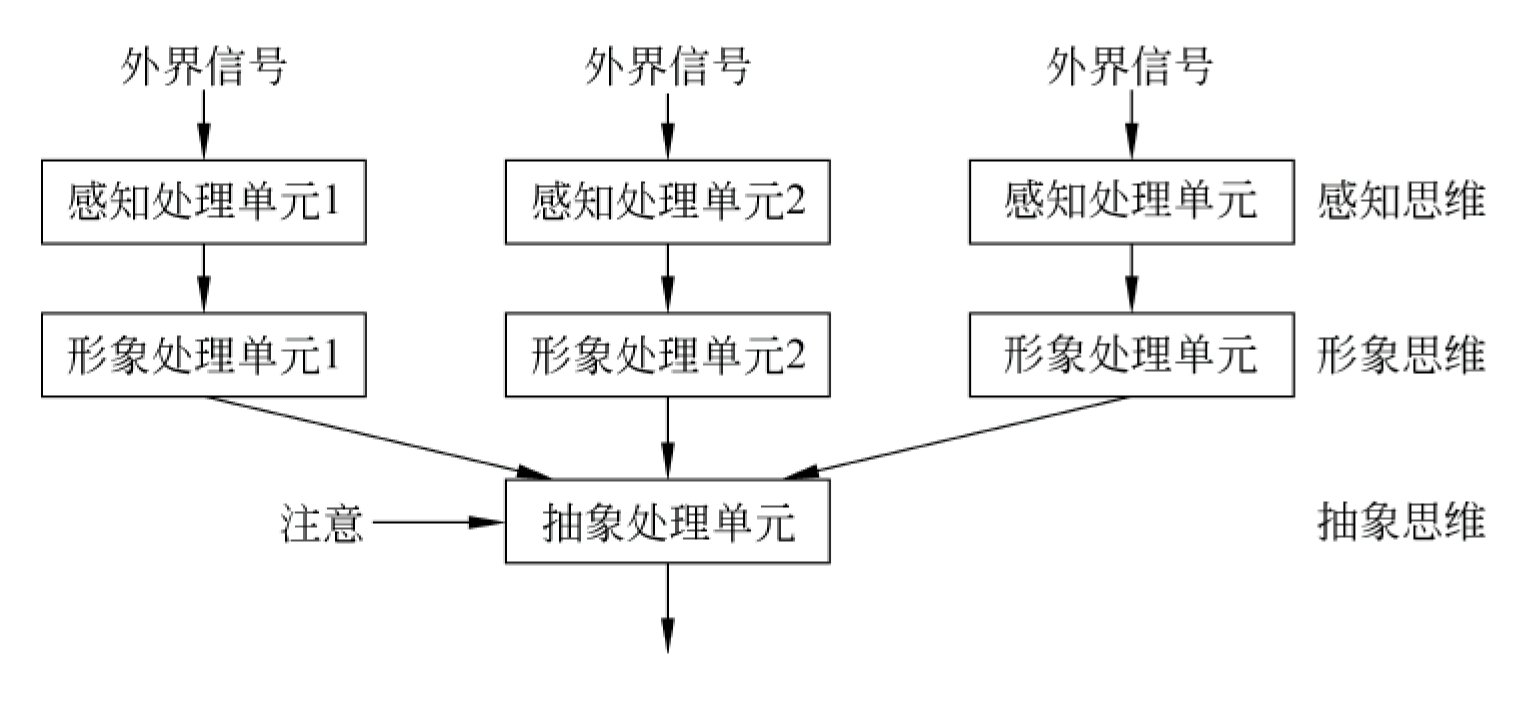



图 10.1 思维的层次模型

思维模型就是要研究这三种思维形式的相互关系,以及它们之间的相互转换的微观过程。人们可以用神经网络的稳定吸引子来表示联想记忆、图像识别的问题。但是要解决从形象思维到逻辑思维的过渡的微过程,还需要作进一步的长期研究。

### 10.2.1 抽象思维

抽象思维凭借科学的抽象概念对事物的本质和客观世界发展的深远过程进行反映,使人们通过认识活动获得远远超出靠感觉器官直接感知的知识。科学的抽象是在概念中反映自然界或社会物质过程的内在本质的思想,它是在对事物的本质属性进行分析、综合、比较的基础上,抽取出事物的本质属性,撇开其非本质属性,使认识从感性的具体进入抽象的规定,形成概念。空洞的、臆造的、不可捉摸的抽象是不科学的抽象。科学的、合乎逻辑的抽象思维是在社会实践的基础上形成的。

抽象思维深刻地反映着外部世界,使人能在认识客观规律的基础上科学地预见事物和现象的发展趋势,预言"生动的直观"没有直接提供出来的、但存在于意识之外的自然现象及其特征。它对科学研究具有重要意义。

在感性认识的基础上,通过概念、判断、推理,反映事物的本质,揭示事物的内部联系的过程是抽象思维。概念是反映事物的本质和内部联系的思维形式。概念不仅是实践的产物,同时也是抽象思维的结果。通过对事物的属性进行分析、综合、比较的基础上,抽取出事物的本质属性,撇开非本质属性,从而形成对某一事物的概念。例如,"人"这个概念,就是在对千差万别的人进行分析、综合、比较的基础上,撇开他们非本质属性(肤色、语言、国别、性别、年龄、职业等),抽取出他们的本质属性(都是能够进行高级思维活动,能够按照一定目的制造和使用工具的动物)而形成的,这就是抽象。概括是指在思想中把某些具有若干相同属性的事物中抽取出来的本质属性,推广到具有这种相同属性的一切事物,从而形成关于这类事物的普遍概念。任何一个科学的概念、范畴和一般原理,都是通过抽象和概括而形成的。一切正确的、科学的抽象和概括所形成的概念和思想,都是更深刻、更全面、更正确地反映着客观事物的本质。

判断是对事物情况有所肯定或否定的思维形式。判断是展开了的概念,它表示概念之间的一定联系和关系。客观事物永远是具体的,因此,要作出恰当的判断,必须注意事物所处的时间、地点和条件。人们的实践和认识是不断发展的,与此相适应,判断的形式也不断变化,从低级到高级,即从单一判断向特殊判断,再向普遍判断转化。

由判断到推理是认识进一步深化的过程。判断是概念之间矛盾的展开,从而使它更深刻地揭露了概念的实质。推理是判断之间矛盾的展开,它揭露了各个判断之间的必然联系,即从已有的判断(前提)逻辑地推论出新的判断(结论)。判断构成推理,在推理中又不断发展。这说明,推理与概念、判断是相互联系,相互促进的。

## 10.2.2 形象思维

形象思维是凭借头脑中储有的表象进行的思维。这种思维活动是右脑进行的,因为右脑主要负责直观的、综合的、几何的、绘画的思考认识和行为。

一个典型的例子是,爱因斯坦这样描述他的思维过程:“我思考问题时,不是用语言进行思考,而是用活动的跳跃的形象进行思考,当这种思考完成以后,我要花很大力气把它们转换成语言。”形象思维或叫直感思维,主要采用典型化的方式进行概括,并用形象材料来思维。形象是形象思维的细胞。形象思维具有以下四个特征。

(1) 形象性,形象材料的最主要特征是形象性,即具体性、直观性。这同抽象思维所使用的概念、理论、数字等显然是不同的。

(2) 概括性。通过典型形象或概括性的形象把握同类事物的共同特征。科学研究中广泛使用的抽样试验、典型病例分析,各种科学模型等,均具有概括性的特点。

(3) 创造性。创造性思维所使用的思维材料和思维产品绝大部分都是加工改造过或重新创造出来的形象。艺术家构思人物形象时和科学家设计新产品时的思维材料都具有这样的特点。既然一切有形物体的创新与改造,一般都表现在形象的变革上,那么设计者在进行这种构思时就必须对思维中的形象加以创造或改造。不仅在创造一个新事物时是如此,而且在用形象思维方式来认识一个现有事物时也不例外。科学家卢瑟福在研究原于内部的结构时,根据粒子散射实验,设想出原子内部像是一个微观的太阳系。原子核居中,电子则在各自的特定轨道上运行,如群星绕日旋转。这便产生了著名的原子行星模型。

(4) 运动性。形象思维作为一种理性认识,它的思维材料不是静止的、孤立的、不变的。提供各种想像、联想与创造性构思,促进思维的运动,对形象进行深入地研究分析,获取所需的知识。

这些特性使形象思维既超出了感性认识进入了理性认识的范围,却又不同于抽象思维,是另一种理性认识。模式识别是典型的形象思维,它用计算机进行模式信息处理,对文字、图像、声音、物体进行分类、描述与分析、理解。目前模式识别已在一定程度上直接或间接地得到应用。已经设计出各种模式信息系统、光学文字识别机、细胞或血球识别机、声音识别装置等,这些在国外已成为商品。模式识别技术也开始用于设计,以利用图像信息为基础的自动检验系统。序列图像分析、计算机视觉、语音理解和图像理解系统的研究与实现已成为普遍感兴趣的方向。

模式识别方法一般分为统计(决策理论)模式识别与句法(结构)模式识别两大类。统计模式识别着眼于找出一组能反应模式特点的特征,首先把模式进行数据压缩以抽取特征,并

考虑到对于通常遇到的干扰和畸变来说,所选的特征具有不变性或者至少是很不敏感。如果抽取 $N$ 个特征能够基本描述原来的模式,那么就用 $N$ 个特征构成的一个向量来代表原来的模式。于是模式识别的统计法就以高维随机向量分析为基础。模式分类就相当于把特征空间划分成若干部分,每一部分与一个模式类相对应。当出现一个新的模式时,就根据描述这个模式的向量位于特征空间的哪一部分而判定属于哪一类;句法模式识别方法,则完全从不同的途径来解决模式识别问题,它着眼于把模式的构成与语言的生成加以类比,借鉴数理语言学的方法与结果。这样,就把识别方法建立在数理语言的基础上。前面介绍过的短语结构文法,稍加修改就可以用来描述和识别图像模式。只要着眼于一幅图像如何由比较简单的子图像构成,子图像又如何由更简单的子图像构成等,就像同英文子句由分句构成,分句又由短语构成,一幅图像就相当于由某种文法规则产生的子句。模式的表达形式可以像语言由符号构成的链那样,是一条由某些特征或基本单元组成的链,也可以是一种树状结构或者是图的形式。识别模式可以体现为描述模式的特征链为某种类型的自动机所接受,或者是对描述模式的特征链(句子)进行句法分析,分析一个句子对于某个文法来说,在句法上是否正确,从而决定这个句子所描述的模式属于哪一类,不仅决定了模式的类别,也给出了一种描述。句法模式识别的研究不像统计模式识别那样透彻,它是最近 15 年才逐渐发展起来的。1974 年美国普渡大学傅京孙教授发表了第二本专著《模式识别中的句法方法》,奠定了句法模式识别的基础[247]。句法方法抓住了图像模式与语言两者之间在结构方面的共性,加以沟通,给模式识别打开了一个新局面。

图像理解是模式识别领域中受到重视和引起兴趣的问题。例如在一幅图像中或者一系列图像中,是否有所要寻找的目标。图像理解也是一个主动的过程,需要利用知识。最终的要求是对图像作出分析与解释。图像理解系统的一般概念如图 10.2 所示。

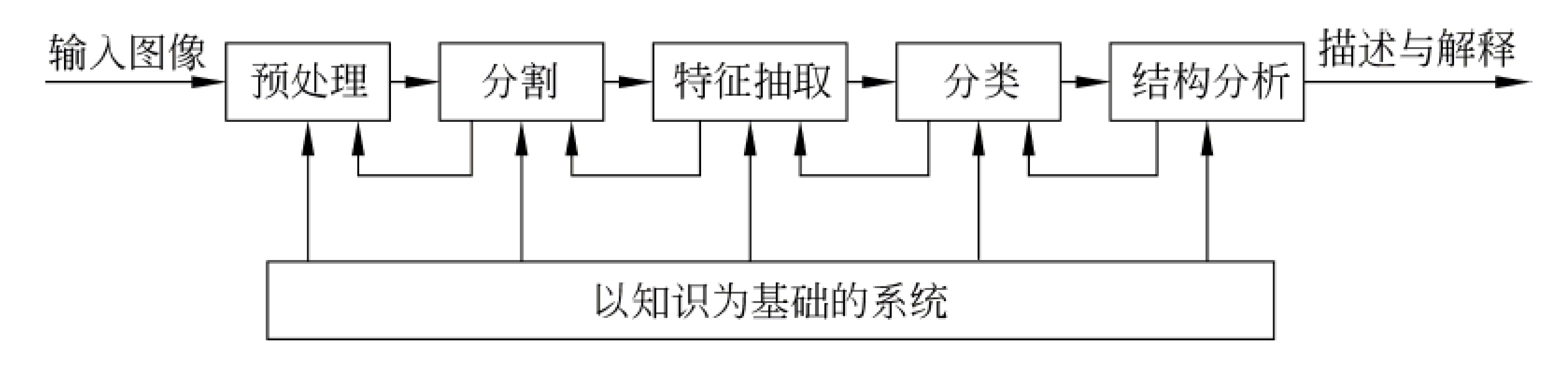



图 10.2　图像理解系统

系统中的预处理,包括图像的增强、复原和编码等。分割是把图像中灰度级不同的部分分开,然后在分开了的各个部分找出能描述该部分性质的特征。特征抽取的目的是为了进行分类;分类的目的是希望对整个图像作结构上的分析;进行结构分析在于给出图像的一种描述,并对图像中的重要信息给予解释。系统中每个环节都对前面的环节有一种反馈作用。在进行图像信息处理与识别的过程中,还需要来自外界的信息。例如来自一个知识库、或者以知识为基础的系统,以知识为基础指的就是人的经验。

平克(Daniel H Pink)在《全新思维》一书中指出:我们的大脑分为两个半球。左半球表示顺序、逻辑和分析能力,右半球则是非线性的、直觉的和整体的[576]。平克认为我们的经济和社会正在从以逻辑、线性、类似计算机的能力为基础的信息时代向概念时代转变,概念时代的经济和社会建立在创造性思维、共情能力和全局能力的基础上。《全新思维》介绍了六种基本的能力,即"六大感知",包括:设计感(Design)、故事感(Story)、交响能力(Symphony)、共情能力(Empathy)、娱乐感(Play)、探寻意义(Meaning)等。

### 10.2.3　灵感思维

灵感思维也称作顿悟。它是人们借助直觉启示所猝然迸发的一种领悟或理解的思维形式。诗人、文学家的“神来之笔”，军事指挥家的“出奇制胜”、思想战略家的“豁然贯通”、科学家、发明家的“茅塞顿开”等，都说明了灵感的这一特点。它是在经过长时间的思索，问题没有得到解决，但是突然受到某一事物的启发，问题一下子解决的思维方法。“十月怀胎，一朝分娩”，就是这种方法的形象化的描写。灵感来自于信息的诱导、经验的积累、联想的升华、事业心的催化。

从 19 世纪后期开始，在西方出现了一股把灵感同人的下意识，特别是同人的自然天性结合起来的反理性主义思潮。唯意志论者叔本华宣称艺术家是“疯人院里的天才。象牙塔里的狂人”；直觉主义者柏格森主张用非理性的直觉去把握世界的本质；而弗洛伊德则把导致艺术灵感的潜意识归结为人的本能欲望，特别是性欲望的冲动。他认为，艺术作品的“精力、它的非理性和它的神秘力量得自本能，我们可以把本能看作是所谓‘灵感’的泉源……从我们称为本能的那片黑暗的心理领域，往往会突如其来地涌现出艺术家用以营造其艺术作品的那些词汇、声音或意象的提示……弗洛伊德为灵感这种情况提供了一种心理学的解释。”

英国《美学》杂志主编奥斯本在 1977 年的夏季号上专文论述了灵感概念在西方的产生和嬗变过程。奥斯本将西方灵感概念的发展过程分为三个阶段：①原始宗教意义上的神赐天启论；②灵感与天才的概念相结合；③灵感与无意识的心理学相结合。他忽略了在西方同时存在一条与唯心主义、神秘主义路线相对应的唯物主义的灵感概念发展路线。

从德谟克利特到狄德罗、到费尔巴哈和车尔尼雪夫斯基，他们在阐述灵感问题时强调的侧面可以不同，但贯穿的认识路线却是一条，就是唯物主义的反映论。这条认识路线的实质，正如马克思和恩格斯所指出的：“观念的东西不外乎是移入人的头脑中改造过的物质的东西而已。”思维着的精神不过是物质“在地球上的最美的花朵”。用这样的观点去解释灵感现象，必然得到这样一个为一切唯物主义者所赞同的结论、灵感，作为人类特有的一种精神现象，只能是地球上最美妙的物质之花，是主观反映客观的认识运动的一种特殊形式。

一般来说，抽象思维发生在显意识，借助于概念实施严格的逻辑推理，从某一前提出发，一步接一步地推论下去，直至得到结论。整个推理过程表现为线性的，一维的。形象思维主要发生在显意识，也时有潜意识参与活动。形象思维，是用形象来思考和表达的。形象思维发生过程，既离不开灵敏、直觉、想像等非逻辑思维的启迪，也少不了按照相似律、对照律等方法的推论；比抽象思维发生过程复杂了，是面型的，二维的。灵感思维主要发生在潜意识，是显意识和潜意识相互交融的结果[89]。灵感的蕴育过程，表现为知觉经验信息、新鲜的课题信息、脑高级神经系统的“建构”活动这三个方面综合进行的拓扑同构而形成的。灵感思维，是非线性的，三维的。因而，灵感思维有着抽象思维、形象思维所不具有的特征。

根据人们的实践经验，诱发灵感的机制大致可分五个阶段，即境域、启迪、跃迁、顿悟、验证。这里分别扼要论述一下。

(1) 境域，是指那种足可诱导灵感迸发的充分且必要的境界。创造性课题在大脑形成

后,必须竭尽全力进入“神动天随,寝食咸废,精凝思极,耳目都融,奇语玄言,恍榴呈露”那样一种精神、心理的全新境界。创造者入境后表现出来的那种潜思维与显思维随意交融,肆意驰骋,神与物游的“忘我”境域,正是“创作的最高境界”。

(2) 启迪,是指机遇诱发灵感的偶然性信息。“万事俱备,只欠东风”。启迪,好比东风,使已有准备的心灵受惠。从认识论来说,启迪是诱导思维发生的一种普遍方式,是连接各种思维信息的纽带,是开启新思路的金钥匙。

(3) 跃迁,是指灵感发生时的那种非逻辑质变方式。显意识和潜意识交互作用,促使潜思维孕育的灵感达到“神思方远,万涂竞萌”之时,正是信息在思维过程中实现跃迁的结果。这种跃迁就是潜思维的特征,是一种跨越推理程序的、非连续的质变方式。

(4) 顿悟,是指灵感在潜意识孕育成熟后,同显意识沟通时的瞬间表现。宋代大哲学家朱熹称顿悟为“豁然贯通”,有时感到“茅塞顿开”等,都说明此时灵感是被意识到了的灵感。

(5) 验证,是指对灵感思维结果的真伪进行科学的分析与鉴定。随灵感的迸发,新概念;新理论、新思路脱颖而出。但是,直觉可能是模糊的,顿悟可能有缺陷。不能认为每一个结论都是有效的,需要进行验证。

人脑是个复杂系统。复杂系统经常采取层级结构构成。层级结构是由相互联系的子系统组成的系统,每个子系统在结构上又是层级式的,直到我们达到某个基本子系统的最低层次。在人的中枢神经系统里是有层次的,而灵感可能是多个自我,是脑子里的不同部分在起作用,忽然接通,问题就解决了。

## 10.3 精神活动层级

明斯基在《情感机器》[507]中指出,情感是人类一种特殊的思维方式,并在洞悉思维本质的基础上,指出人类思维的运行方式,提出了塑造未来机器的 6 大维度:意识、精神活动、常识、思维、智能和自我。我们的大脑是如何产生如此多新事物和新想法的?资源可以分为 6 种不同的层级——本能反应、后天反应、沉思、反思、自我反思和自我意识反思,以对想法和思维机制进行衡量(见图 10.3)[507]。每一个层级模式都建立在下一个层级模式的基础之上,最上层的模式表现的是人们的最高理想和个人目标。



图 10.3 精神活动的 6 大层级

### 1. 本能反应

我们天生就拥有本能反应,它会保护我们获得生存能力。在 20 世纪的心理学领域,“刺激-反应”模型极受欢迎,有些研究甚至认为这种模型能够解释人类的所有行为。这种“刺激-反应”模型就是 If→Do 规则。由于大多数行为的发生取决于我们自身所处的环境,这些简单的规则很少会起作用,例如,“如果看见食物,就吃掉它”规则会迫使你吃掉自己看到的所有食物,不管你是否感到饥饿或是否需要食物。为了防止此类情形的出现,每一个 If 必须指明具体的目标。明斯基提出了更强大的规则,即 If+Do→Then 规则。

### 2. 后天反应

认识到一些情况需要特殊的应对方式。所有动物具有与生俱来的“远离逼近物体”的本能。只要动物一直生活在激发本能的环境中，这些固有本能就能有效地发挥作用；但是，一旦动物生活的世界发生变化，每一物种可能都需要学习新的反应方式。

当遭遇新情况时，动物会随机应变地采取一些行动，而如果在某些行为中尝到了“甜头”，这些行为便会在动物的大脑里得到“强化”，因此当动物再次遭遇相同的情况时，就更可能重复之前的这些行为。

天生、本能的反应左右着我们对很多事情的反馈，但我们也在不断更新着新的反应方法，这需要人类大脑组织模型的第二个层次。

### 3. 沉思

为了实现更为复杂的目标，我们就需要通过使用从过去经历中学到的全部知识来制定更为详细的计划。这种内部精神活动赋予了人类很多特殊的能力。

如果你处在情况 A 中，且想进入情况 Z，那你就可能已经知道实现这种想法的规则了，例如 If 情况 A→Do 行动→Then 情况 Z。这种情况下仅执行行动就能实现目标。

但如果你根本不知道这类规则的存在呢？那时你便可能搜索记忆，寻找两个规则链，通过中间情况 M 来实现目标。

If 情况 A→Do 行动 1→Then 情况 M，然后

If 情况 M→Do 行动 2→Then 情况 Z

如果一个或两个这样的步骤不能解决问题时，必须搜索寻找更多的步骤，尽可能缩小搜索范围。

### 4. 反思

反思自己的决定。不是对外部活动，而是对大脑活动作出反应。当唱一支娴熟的歌曲时，在开始前，期望集中于整个歌曲；开唱后，从期望进入记忆，因此活动向两面展开：已经唱出的属于记忆，未唱的属于期望；注意力则在当下，要将未来引来，并将它变为过去。

然而，为结合描述进行推理，每一层次都需要短时记忆、假设和结论。大脑应该反思何种思维活动？应该包括以下活动：错误的预测、受阻的计划和无法获取的知识。机器并不像“自我意识实体”一样可以对自我进行全面的认识，但一旦机器拥有表现自身广泛活动的多个模型，便可以实现这种认识。有时，模型有助于系统中的一些部位思考其他部位发生的事情，但是，如果想让系统思考其本身的所有细节则是不切实际的。

### 5. 自我反思

人类不同于动物的另一种能力是，人类像一个思想家一样拥有自我意识和反思能力，然而动物从不像思想家一样反思自我，因为从思想的实质方面说，动物从来无法清楚地区分思想本身和思想机制。没有任何自我模型是完整的，最好的做法是同时创建几个模型，每个模型描述其中一个部分。

认识到自我反思的重要性，就如同认识到自身的困惑一样，是件颇具智慧的事情。因为

只有当我们意识到自我困惑时,才会知道该升华动机和目标了。自我反思有助于我们认识到以下几个问题:不知道自己要做什么,或在更为无关痛痒的细节上浪费时间,或正在追寻一个不太合适的目标。自我反思也有助于我们制订更好的计划、思考范围更大的情感活动。人们何时会使用自己的高层次思维方式?当人们不能使用常规思维系统时,反思性思维方式便开始起作用。

#### 6. 自我意识反思

出生之后学习到的自我反应能力被称为后天反应。沉思和反思有助于解决更复杂的难题。有些问题牵涉到自我模型、未来的可能结果时,便进入自我反思的研究范围。当你反思自己近期的想法时,你会思考这种反思本身吗?与其相似,自我反思也是反思的一种,想知道自己的行为在多大程度上实现了自己的理想。

这些区别并不是很明显,即使我们最简单的想法都会涉及所谓对分配时间和资源的“反思”。那么“如果不能使用一种方法,我将尝试另一种方法”或“我已经在那件事上花费了很长时间”。最后,大脑变成是一个异常复杂的系统,任何单一的模型都不能够解读大脑,除非这个模型本身相当复杂,但那时这一模型也会因为太过复杂而变得毫无用处。因此,心理学家们需要成倍地扩充思维(和大脑)模式,每一种模式都可以解释思维的不同方面和类型,尤其是在个体处理经济、宗教和种族问题时可能会相互矛盾时,人类自我意识反思的模式更应增强。

## 10.4 推理

推理是从已有的知识得出新知识的思维形式,在推理中可以清楚地看到人类思维的创造性。人们用推理的方法去认识那些本能直接观察到的现实过程时,只要能够在实践中证实推理的复杂链条中的必需的重要环节,并且合乎逻辑地进行推论,那么所做出的新判断(结论)、提出的新概念就是科学的。

作为心理学最古老的研究领域之一,推理研究表明人是理性的,思维的原则就是逻辑原则。这一基本思想已经被应用到了推理心理学的研究中。推理心理学的研究主要包括两方面的内容:演绎推理和归纳推理。当人们进行演绎推理时,通常是假设在某些表述或者前提成立的条件下推测必然会出现什么结果。在进行归纳推理时,人们从描述个别例子的前提中概括出一般结论。下面扼要介绍演绎推理、归纳推理、反绎推理、类比推理、非单调推理,以及常理性推理。

### 10.4.1 演绎推理

演绎推理是前提与结论之间有蕴涵关系的推理,或者说,前提与结论之间有必然联系的推理,由一般推演出特殊。

$A$ 是 $\Gamma$(即 $\Gamma$ 中公式)的逻辑推理,记作 $\Gamma \models A$,当且仅当任何不空论域中的任何赋值 $\varphi$,如果 $\varphi(\Gamma)=1$,则 $\varphi(A)=1$。

给定不空论域 $S$。当 $S$ 中任何赋值 $\varphi$ 都使得

$$\varphi(\Gamma) \Rightarrow 1 \Rightarrow \varphi(A) = 1$$

成立时，我们说在 $S$ 中 $A$ 是 $\Gamma$ 的逻辑推理，记作：在 $S$ 中 $\Gamma \models A$。

数理逻辑中研究推理，通过形式推理系统研究演绎推理。可以证明，凡是形式推理所反映的前提与结论之间的关系在演绎推理中都是成立的，因此形式推理没有超出演绎推理的范围，形式推理可靠地反映了演绎推理；凡是在演绎推理中成立的前提与结论之间的关系，形式推理都是能反映的，因此形式推理在反映演绎推理时并没有遗漏，形式推理对于反映演绎推理来说是完备的。

经常应用的一种推理形式是三段论。它由且只由三个性质判断组成，其中两个性质判断是前提，另一性质判断是结论。就主项和谓项说，它包含而且只包含三个不同的概念，每个概念在两个判断中各出现一次。这三个不同的概念，分别叫做大项、小项与中项。大项是作为结论的谓项的那个概念，用 P 表示。小项是作为结论的主项的那个概念，用 S 表示。中项是在两个前提中都出现的那个概念，用 M 表示。由于大项、中项与小项在前提中位置不同而形成的各种不同的三段论形式，叫做三段论的格。三段论有下面四个格：

第一格：　M—P
　　　　　S—M
　　　　　——
　　　　　S—P

第二格：　P—M
　　　　　S—M
　　　　　——
　　　　　S—P

第三格：　M—P
　　　　　M—S
　　　　　——
　　　　　S—P

第四格：　P—M
　　　　　M—S
　　　　　——
　　　　　S—P

基于规则的演绎系统中，一般分为正向系统、逆向系统和双向综合系统。在基于规则的正向演绎系统中，作为 F 规则用的蕴涵式对事实的总数据库进行操作运算，直至得到该目标公式的一个终止条件为止。在基于规则的逆向演绎系统中，作为 B 规则用的蕴涵式对目标的总数据库进行操作运算，直至得到包含这些事实的终止条件为止。在基于规则的正向逆向综合系统中，分别从两个方向应用不同的规则（F 规则或 B 规则）进行操作运算。这种系统是一种直接证明系统，而不是归结反演系统。

## 10.4.2　归纳推理

归纳推理是由个别的事物或现象推出该类事物或现象的普遍性规律的推理，这种推理反映前提与结论之间有或然性联系。这是由特殊推出一般的思维过程。

近数十年来国外归纳逻辑的研究在两个方向上进行着,一是在培根归纳逻辑的古典意义上继续寻找从经验事实导出相应的普遍原理的逻辑途径;另一个方向是运用概率论和形式化、合理化的手段来探索有限经验事实对适应于一定范围的普遍命题的“支持”或“确证”的程度,这种逻辑实际上是一种理论评价的逻辑。

培根关于归纳法的主要思想如下。

(1) 感官必须得到帮助和指导以克服感性认识的片面性和表面性。

(2) 构成概念时要经过适当的归纳程序。

(3) 公理的构成应当用逐步上升的方法。

(4) 必须重视反演法和排斥法等在归纳过程中的作用。

这些思想构成了后来穆勒(John S Mill)提出的归纳法四条规则的基础。穆勒在《逻辑体系》一书中提出了有关归纳的契合法、差异法、共变法和剩余法。他认为“归纳可定义为发现和证明一般命题的操作”。

概率逻辑是 20 世纪 30 年代兴起的。莱欣巴哈以相对频率为基础,利用概率论的数学工具来求出一个命题的频率极限,并以此来预测未来事件。21 世纪四五十年代,卡尔纳普建立以合理信念为基础的概率逻辑。他采取贝叶斯主义的立场,把合理信念直接地描绘成概率函数,并把概率看作代表一个陈述和另一个证据陈述之间的逻辑关系。贝叶斯定理可表示为:

$$P(h \mid e) = P(e \mid h)\frac{P(h)}{P(e)} \tag{10.2}$$

其中 $P$ 为概率,$h$ 为假设,$e$ 为证据。这个公式就是说 $h$ 相对于 $e$ 的概率等于 $h$ 相对于 $e$ 的似然值,也就是如果 $h$ 为真 $e$ 的概率乘以 $h$ 的先验概念与 $e$ 的先验概率之比。这里先验概率是指在这次试验前已经知道的概率。如 $A_1, A_2, \cdots$,是导致试验结果的“原因”,$P(A_i)$ 就称先验概率。若试验产生了事件 $B$,这个信息将有助于探讨事件发生的“原因”。条件概率 $P(A_i|B)$ 就称为后验概率。所以,卡尔纳普在采取了贝叶斯的立场之后,就要对先验概率作出合理的解释。卡尔纳普不同意把先验概率仅仅理解为个人的主观的相信度,而力求对合理信念作出比较客观的解释。

卡尔纳普把他的概率的逻辑概念理解为“确定程度”。他用符号 $C$ 表示他所理解的概率概念,用“$C(h, e)$”表示“假设 $h$ 相对于证据 $e$ 的确定程度”,并进一步引入可信任函项 $Cred$ 和信念函项 $Cr$ 来解决怎样将归纳逻辑应用于合理的决策。确定度 $C$ 定义为

$$C(h, e_1, e_2, \cdots, e_n) = r \tag{10.3}$$

即陈述(归纳前提)$e_1, e_2, \cdots, e_n$ 联合起来将逻辑概率 $r$ 给予陈述(归纳结论)$h$。这样,卡尔纳普又依据某人 $x$ 在 $t$ 时对某一条件概率所寄予的价值的期望定义了信念函项等概念。他把假设陈述 $H$ 相对于证据陈述 $E$ 的信念函项 $Cr$ 定义为:

$$C_{rx,t}(H/E) = \frac{Crx, t(E \cap H)}{C_{rx,t}(E)} \tag{10.4}$$

进一步,他定义了可信任函项 $Cred$,某观察者 $x$,在 $T$ 时所有的观察知识是 $A$,则他在 $T$ 时对 $H$ 的信任程度是 $Cred(H/A)$。信念函项是以可信任函项为基础的,即

$$C_{rT}(H_1) = Cred(H_1/A_1) \tag{10.5}$$

卡尔纳普认为有了这两个概念,我们就能从规范的决策理论过渡到归纳逻辑,并将信念

函项和可信任函项与纯逻辑概念相对应。相应 $Cr$ 他称为 $m$-函项，即归纳量度画项。相应于 $Cred$ 他称为 $C$-函项，即归纳确证函项。这样就可把概率演算的通用公理作为对 $m$ 的归纳逻辑的基本公理。他指出："在归纳逻辑中，$C$-函项较 $m$-函项更重要，因为某一 $C$ 值表示信念的合理程度并能帮助在合理决策中作出决定。$m$-函项则是主要作为定义 $C$-函项并决定其值的方便手段。$m$-函项在公理的概率演算的意义上公种(绝对的)概率函项；而 $C$-函项则是一个条件的(相对的)概率函项。"他说如果我们把 $C$ 作为原始词项，则对 $C$ 的公理可以这样来阐述：

(1) 下限公理：$C(H/E)\geqslant 0$ (10.6)

(2) 自我确证公理：$C(H/E)=1$ (10.7)

(3) 互补公理：$C(H/E)+C(-H/E)=1$ (10.8)

(4) 一般乘法公理：若 $E\cap H$ 是可能的

$$C(H\cap H'/E)=C(H/E)C(H'/E\cap H) \tag{10.9}$$

卡尔纳普正是在这些公理的基础上来构建他的归纳逻辑的形式化体系的。

关于归纳逻辑的合理性是哲学史上长期争论的问题。休谟提出归纳疑难，其核心思想就是不能依据过去而推断未来，不能依据个别而推断一般。休谟认为"根据经验来的一切推论都是习惯的结果，而不是理性的结果"。从而就可以得出归纳法的合理性是不可证明的，与之相联系的经验科学也没有合理性的不可知论的结论。

波普尔在《客观知识》一书中，将归纳疑难表达为：

(1) 归纳能否被证明？

(2) 归纳原理能否被证明？

(3) 能否证明这样一些归纳原理，如"未来与过去一样"或证明所谓的"自然齐一律"。

### 10.4.3　反绎推理

反绎推理(Abduction)也称为溯因推理。在反绎推理中，我们给定规则 $p\Rightarrow q$ 和 $q$ 的合理信念。然后希望在某种解释下得到谓词 $p$ 为真。

基于逻辑的办法则是建立在解释的更高级的概念的基础上。莱维斯克(Levesque)于1989年定义某些前面无法解释的现象集合 $\boldsymbol{O}$ 为假设集 $\boldsymbol{H}$ 中与背景知识 $\boldsymbol{K}$ 的最小集合[427]。假设 $\boldsymbol{H}$ 连同背景知识 $\boldsymbol{K}$ 必须能推导解释出 $\boldsymbol{O}$。更形式化一点：

$$\mathbf{abduce}(\boldsymbol{K},\boldsymbol{O})=\boldsymbol{H}$$

当且仅当

(1) $\boldsymbol{K}$ 不能推导解释(entail)出 $\boldsymbol{O}$；

(2) $\boldsymbol{H}\cup\boldsymbol{K}$ 能推导解释出 $\boldsymbol{O}$；

(3) $\boldsymbol{H}\cup\boldsymbol{K}$ 是一致的；

(4) $\boldsymbol{H}\cup\boldsymbol{K}$ 不存在 $\boldsymbol{H}$ 的子集有性质 1、2 和 3。

需要指出的是总的来说，可能会存在许多假设集，也就是说，对一个给定的现象可能会有很多潜在的解释集。

基于逻辑的反绎解释的定义暗示了发现知识库系统中的内容的解释有相应的机制。如

果可解释的假设必须能推导解释出现象 **O**,则建立一个完整的解释的方式就是从 **O** 向后推理。

### 10.4.4 类比推理

依据两个对象之间存在着某种类似或相似的关系,从已知这一对象有某种性质而推出另一对象具有某一相应的性质的推理过程称为类比推理。类比推理的模式可以表示为:

如果　A 有属性 abcd
　　　B 有属性 abc
则　　B 可能有属性 d

所以,类比推理是依据早先获得的关于某一系统的知识,作为推测另一类似的系统的信息的手段。这种类比推理的客观基础在于事物、过程和系统之间各要素的普遍联系,以及这种联系之间所存在着的可比较的客观基础。

当我们把解决某个问题取得的经验用来解决类似的问题时,关键就要善于发现不同问题之间的类似的地方。波兰数学家巴拿赫曾说:"一个人是数学家,那是因为他善于发现判断之间的类似;如果他能判明论证之间的类似,他就是个优秀的数学家;要是他能识破理论之间的类似,那么,他就成了杰出的数学家。可是,我认为还应当有这样的数学家,他能够洞察类似之间的类似。"不仅数学如此,物理学、化学、生物学、天文学、遗传学、地学也是如此。不仅学科内部有相似性,学科之间也有相似性。认知科学正是在人脑的信息加工与计算机处理相似的基础上建立起来的。在社会科学、文学、艺术、音乐等也离不开相似。相似就是客观事物的存在的同与变异矛盾的统一。变异就是事物发展过程中的差异。根据相似的因素,可以把相似现象分为功能相似、结构相似、动力相似、几何相似等。邦格(M Bunge)在《方法、模型和物质》一文中将类比进行分类,如表 10.1 所示。其中 $N$ 表示事物,$A$ 表示人工事物,$C$ 表示结构,≈表示相似。

类比推理的原理如图 10.4 所示,图中 $\beta_i$ 和 $\alpha$ 是 $S_i$ 成立的事实。$\beta_i'$是 $S_2$ 中成立的事实。$\varphi$ 是对象之间关系的相似性;所谓类比推理,即 $\beta_i\varphi\beta'i(1\leqslant i\leqslant n)$时,$\alpha\varphi\alpha'$的 $S_2$ 中。将推论得出 $\alpha'$。为了实现这样的类比推理,给出下列条件是必要的。

对象　$S_1$:前提 $\beta_1,\cdots,\beta_n\rightarrow$结论 $\alpha$
相似性 $\varphi$
对象　$S_2$:前提 $\beta_1',\cdots,\beta_n'\rightarrow$结论 $\alpha'$?

图 10.4　类比推理的原理

① 相似性 $\varphi$ 的定义。

② 从所给对象 $S_1$ 和 $S_2$ 求出 $\varphi$ 的方法。

③ 为了推出 $\alpha\varphi\alpha'$的 $\alpha'$的操作。

首先,相似性 $\varphi$ 的定义应该与对象的表现和它的意义有关。这里,类推的对象是判断子句的有限集合 $S_i$。定义子句 $A,\beta_i$ 作为文字常量,构成"if-then 规则"。不含 $S_i$ 的变元的项 $t_i$,表示对象的个体,谓词符号 $P$ 表示个体间的关系,不含变元的原子逻辑式 $P(t_i,\cdots,t_n)$表示个体 $t_i$ 间的关系 $P$。以后,不含变元的原子逻辑式简称为原子。由全部原子构成的集合用 $\beta_i$ 表示。$S_i$ 的最小模型 $M_i$ 是由 $S_i$ 逻辑地推出的原子的集合。

**表 10.1　邦格的类比分类**

| 符　　号 | 种　　类 | 例　　子 |
|---|---|---|
| $\approx_1 \subset N\times N$ | 物——物相似 | 有机体——社会 |
| $\approx_2 \subset N\times A$ | 物——人工物相似 | 细胞——工厂 |
| $\approx_3 \subset N\times C$ | 物——结构相似 | 有机体——理论 |
| $\approx_4 \subset A\times N$ | 人工物——物相似(与$\approx_2$同) | 电子枪——枪 |
| $\approx_5 \subset A\times A$ | 人工物——人工物相似 | |
| $\approx_6 \subset A\times C$ | 人工物——结构相似 | |
| $\approx_7 \subset C\times N$ | 结构——物相似(与$\approx_3$同) | 计算机——自动机理论 |
| $\approx_8 \subset C\times A$ | 结构——人工物相似(与$\approx_6$同) | |
| $\approx_9 \subset C\times C$ | 结构——结构相似 | 理论 * ——理论 |

$$\boldsymbol{M}_i = \{\alpha \in \beta_i : S_i \vdash \alpha\} \tag{10.10}$$

这里，$\vdash$在谓词逻辑中表示逻辑推理，并且，将 $\boldsymbol{M}_i$ 的个体称为 $\boldsymbol{S}_i$ 的事实。

为了确定 $\boldsymbol{M}_i$ 间的类比关系，考虑 $\boldsymbol{M}_i$ 的对应关系是必要的。为此，给 $\boldsymbol{S}_i$ 的项的对应定义如下：

(1) $\boldsymbol{U}(\boldsymbol{S}_i)$作为 $\boldsymbol{S}_i$ 不含变元的集合，这时，$\varphi\subseteq \boldsymbol{U}(\boldsymbol{S}_1)\times \boldsymbol{U}(\boldsymbol{S}_2)$的有限 $\varphi$ 称为配对，而$\times$表示集合的直积。

因为项表示对象 $\boldsymbol{S}_i$ 的个体，配对可以解释为表示个体间某种对应关系。$\boldsymbol{S}_i$ 一般不是常数，而具有函数符号，因此对应对 $\varphi$ 可以扩充为 $\boldsymbol{U}(\boldsymbol{S}_i)$间对应关系 $\varphi^+$。

(2) 由 $\varphi$ 生成的项的关系 $\varphi^+$，若满足下列关系，则定义为最小关系：

① $\varphi \subseteq \varphi^+$　(10.11)

② $<t_i, t_i'> E\varphi^+ (1\leqslant i\leqslant n) \Rightarrow < f(t_1,\cdots,t_n), f(t_1',\cdots,t_n') \in \varphi^+$　(10.12)

其中，$\boldsymbol{f}$ 是 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 共同表示的函数符号。

可是，对于所给的 $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$，可能的配对一般存在多个。各配对 $\varphi$ 每个关系一致，即谓词符号一致的类比可以按下面方法确定。

(3) 设 $\boldsymbol{t}_j \in \boldsymbol{U}(\boldsymbol{S}_i)$，$\varphi$ 是配对，$\alpha,\alpha'$分别是 $\boldsymbol{S}_1, \boldsymbol{S}_2$ 的事实，这时 $\alpha$ 和 $\alpha'$由 $\varphi$ 看成相同，是指对于谓词符号 $P$，写成

$$\alpha = P(t_1,\cdots,t_n)$$
$$\alpha' = P(t_1',\cdots,t_n')$$

并且$<t_j, t_j'> \in \varphi^+$。依据 $\varphi$，$\alpha$ 和 $\alpha'$看成相同，将写成 $\alpha\varphi\alpha'$。根据类推，为了求得原子 $\alpha'$，首先将规则

$$\boldsymbol{R}' : \alpha' \leftarrow \beta_1', \cdots, \beta_n'$$

利用 $\varphi$ 由规则

$$\boldsymbol{R} : \alpha \leftarrow \beta_1, \cdots, \beta_n$$

作出。由 $\alpha' \leftarrow \beta_1', \cdots, \beta_n'$和已知事实$\beta_1', \cdots, \beta_n'$，经用三段论就可推出 $\alpha'$。基于 $\varphi$，由 $\boldsymbol{R}$ 构造 $\boldsymbol{R}'$称为规则的变换，求解 $\alpha'$的过程可以表达成图 10.5 的基本图式。

在基本图式中，例示和三段论法是属演绎推理。如果规则变换可以在演绎系统内实现，那么类比推理就可以在演绎系统中统一处理。

| 例示 | $A \leftarrow B_1, \cdots, B_n$ |
| --- | --- |
| 利用 $\varphi$ 进行规则变换 | $\alpha' \leftarrow \beta_1', \cdots, \beta_n'$ |
| 三段论法 | $\dfrac{\beta_1', \cdots, \beta_n' \quad \alpha' \leftarrow \beta_1', \cdots, \beta_n'}{\alpha'}$ |

图 10.5 基本图式

## 10.4.5 非单调推理

经典逻辑,如形式逻辑、演绎逻辑等,对人类认识世界的处理是单调的。设 $\boldsymbol{A}$ 表示推理规则集,则单调逻辑的语言 $\mathrm{Th}(\boldsymbol{S})=\{\boldsymbol{P} \mid \boldsymbol{S} \rightarrow \boldsymbol{P}\}$ 具有如下单调性。

(1) $\boldsymbol{A} \subseteq \mathrm{Th}(\boldsymbol{A})$

(2) 如果 $\boldsymbol{A} \subseteq \boldsymbol{S}$,则 $\mathrm{Th}(\boldsymbol{A}) \subseteq \mathrm{Th}(\boldsymbol{S})$

(3) $\mathrm{Th}(\mathrm{Th}(\boldsymbol{A}))=\mathrm{Th}(\boldsymbol{A})$

单调推理规则的显著特征之一就是它的语言是封闭的最小固定点,亦即

$$\mathrm{Th}(\boldsymbol{A}) = \bigcap \{s \mid \boldsymbol{A} \rightarrow \boldsymbol{S} \text{ 且 } \mathrm{Th}(\boldsymbol{A}) = \boldsymbol{S}\}$$

具体来说,设有知识系统 $A$,如果已知 $A$ 蕴涵着知识 $\boldsymbol{S}$,即 $A \rightarrow \boldsymbol{S}$,则就可推理得出知识 $\boldsymbol{S}$。但是,在这种基础上建立起来的推理系统与人类对客观世界的认识过程往往是不一致的。基于经典逻辑的推理是人们推理的理想化模型,在日常生活中或是在某些人工智能应用系统中,人们经常要依据某些一般来说是正确的但并非绝对正确的规则进行推理,或者在信息不完全的情况下进行推理,这种推理所得的结论是暂时的,可能会修改的,因而不具有单调性,因此人们称之为非单调推理。

在一阶逻辑中,我们用 $\forall x \boldsymbol{P}(x)=1$ 表示“所有 $x$ 都具有性质 $\boldsymbol{P}$”这一事实。可是实际生活中,这类句子都是近于真实而不是绝对正确的,即大多数 $x$ 具有性质 $\boldsymbol{P}$,但偶然也可能会遇见某些例外。例如,所有的鸟儿都能飞,但企鹅和鸵鸟等例外。所有的桔子是黄的,但未熟的和变异的品种例外。由于这类综合性概括语句不是绝对正确的,采用这些语句进行的推理也不可避免地要产生错误。解决这个问题的一种办法是完全抛弃这类语句,这样虽然不会产生错误,但同时也失去了近于真实的东西和许多本来可以得到的结论。另一种办法是修改这类语句,待它完全正确时再使用,可是这种修改相当困难,即使修改好了,句子的结构已变得相当复杂,无法灵活地使用。一种简便而又妥善的处理办法就是先假定这类语句是正确的,并依据它们进行推理,如果在获取了新的事实后发现原来的结论有问题,推理就具有了非单调性。

为了使非单调推理得到强有力的逻辑支持,人们提出了各种不同的非单调逻辑。其中较为著名的工作有 R Reiter 的默认逻辑,J McCarthy 的限制逻辑, Doyle 的真值维护系统以及 R C Moore 的自认知逻辑等等。

### 1. 默认逻辑

基本思想是在推理过程中,一些真假不能确定而又必须确定的命题,如果假定为真,不产生矛盾,则默认这些命题成立。

一个默认理论定义为二元式($\boldsymbol{D}$,$\boldsymbol{W}$),其中 $\boldsymbol{D}$ 是默认集,$\boldsymbol{W}$ 是封闭的合适公式集。默认规则的一般形式表示为:

$$\frac{\alpha(X): M\beta_1(X), \cdots, M\beta_m(X)}{W(X)} \tag{10.13}$$

其中，$\alpha(X)$，$\beta_1(X)$，…，$\beta_m(X)$都是合适公式，自由变量处在 $X=x_1,\cdots,x_n$ 之中。$\alpha(X)$称为前件，$W(X)$称为结论。如果 $\alpha,\beta_1,\cdots,\beta_m$，$\boldsymbol{W}$ 不包含自由变量，就称默认是封闭的。

### 2. 限定逻辑

限定逻辑及其推理系统是 McCarthy 最早提出的。最初的限定推理主要处理的是“谓词限定”，其基本方法是极小化特定谓词的扩展。在限定逻辑中只有当证明事物满足性质 $P$ 时，才认为它具有性质 $\boldsymbol{P}$。

### 3. 真值维护系统(Truth Maintenance System)

Doyle 的真值维护系统是一个已经实现了的非单调推理系统。它用以协助其他推理程序维护系统的正确性。它的作用并不是生成新的推理，而是在其他程序所产生的命题之间保持相容性。一旦发现某个不相容，它就调出自己的推理机制，面向从属关系的回溯，并通过修改最小的信任集来消除不相容。

## 10.4.6 常识性推理

常识性推理是人工智能的一个领域，旨在帮助计算机更自然地理解人的意思以及跟人进行交互，其方式是收集所有背景假设，并将它们教给计算机。常识性推理具有代表性的系统是 Cycrop 公司的 Cyc 系统，它运营着一个基于逻辑的常识知识库。

Cyc 系统是由莱纳特(Doug Lenat)在 1984 年开始研制的[412]。该项目最开始的目标是将上百万条知识编码成机器可用的形式，用以表示人类常识。CycL 是 Cyc 项目专有的知识表示语言，这种知识表示语言是基于一阶关系的。1986 年莱纳特预测如果想要完成 Cyc 这样庞大的常识知识系统，这将涉及 25 万条规则，并将要花费 350 个人年才能完成。1994 年，Cyc 项目从该公司独立出去，并以此为基础，在美国得克萨斯州奥斯丁成立了 Cycorp 公司。

2009 年 7 月发布了 OpenCyc 2.0 版，涵盖了完整的 Cyc 本体，其中包含了 47000 个概念、306000 个事实，主要是分类断言，并不包含 Cyc 中的复杂规则。这些资源都采取 CycL 语言来进行描述，该语言采取谓词代数描述，语法上与 Lisp 程序设计语言类似。CycL 和 SubL 解释器(允许用户浏览并编辑知识库、具有推理功能)是免费发布给用户的，但是仅包含二进制文件，并不包含源代码。OpenCyc 具有针对 Linux 操作系统及微软 Windows 操作系统的发行版。开源项目 Texai 项目发布了 RDF 版本的 OpenCyc 知识库。

Cyc 中的概念被称为“常量(Constants)”。常量以“#＄”开头并区分大小写。常量主要分为以下几类。

(1) 个体(Individuals)：例如#＄BillClinton 又如#＄France。

(2) 集合(Collections)：例如#＄Tree-ThePlant(包含所有的树)，又如#＄EquivalenceRelation(包含所有的等价关系)。集合中的个体被称为该集合的实例(Instance)。

(3) 真值函数(Truth Functions)：该函数可被应用于一个或多个概念，并返回“真”或

“假”。例如#$siblings表示兄弟姐妹关系,若两个参数对应的内容为兄弟姐妹关系,则该概念返回真值。约定真值函数以小写字母开头,并且可以被拆分为若干个逻辑连接词(例如#$and、#$or、#$not、#$implies)、量词(#$forAll,#$thereExists等)以及谓词。

(4) 函数(Functions):用于以现有术语为基础产生新的术语。例如#$FruitFn具有以下作用:若接收到用于描述一种(或一个集合)植物的声明,则会返回其果实。约定函数常量以大写字母开头,并以“Fn”作为结尾。

Cyc中的谓词最重要的是#$isa以及#$genls。#$isa表示某个对象是某集合的实例,#$genls表示某个集合是另外一个集合的子集合。由概念构成的事实采用CycL语言描述的“句子”表示。谓词则写在与其相关的对象之前,并以括号括起来:

(#$isa #$BillClinton #$UnitedStatesPresident)表示“Bill Clinton属于美国总统集合”;

(#$genls #$Tree-ThePlant #$Plant)表示“所有的树都是植物”;

(#$capitalCity #$France #$Paris)表示“巴黎是法国的首都”。

句子中可以包含变量,变量字符串以“?”开头,这些句子被称为“规则”。与#$isa谓词有关的一条规则如下所示:

(#$isa ?OBJ ?SUBSET)

(#$genls ?SUBSET ?SUPERSET)

(#$isa ?OBJ ?SUPERSET)

上面的规则可解释为:若OBJ为集合SUBSET中的一个实例,并且SUBSET是SUPERSET的子集,则OBJ是集合SUPERSET的一个实例。下面再给出另外一个典型的示例:

(#$relationAllExists #$biologicalMother #$ChordataPhylum #$FemaleAnimal)上面的规则可解释为:对于脊索动物(Chordate)集合#$ChordataPhylum中的所有实例,都存在一个母性动物(为#$FemaleAnimal的实例)作为其母亲(通过谓词#$biologicalMother描述)。

Cyc知识库是由许多microtheories(Mt)构成的,概念集合和事实集合一般与特定的Mt关联。与整体的知识库有所不同的是,每一个Mt相互之间并不矛盾,每一个Mt具有一个常量名,Mt常量约定以字符串“Mt”结尾。例如:#$MathMt表示包含数学知识的Mt,Mt之间可以相互继承得到并组织成一个层次化的结构。例如#$MathMt特化到更为精细的层次便包含了如#$GeometryGMt,即有关几何的Mt。

Cyc推理引擎是从知识库中经过推理获取答案的计算机程序。Cyc推理引擎支持一般的逻辑演绎推理,包括肯定前件假言推理(Modus Ponens)、否定后件假言推理(Modus Tollens)、全称量化(Universal Quantification)、存在量化(Existential Quantification)。

2011年2月14日,IBM公司“沃森”超级计算机(Watson)在美国著名老牌智力游戏节目《危险边缘》(*Jeopardy!*),与肯·詹宁斯和布拉德·鲁特尔比赛。2月16日,经历三轮比赛,智能计算机沃森(Watson)最终赢得问答节目《危险边缘》的冠军,勇夺100万美元大奖。智能计算机沃森成功地采用了常识性推理Cyc系统。

## 10.5　问题求解

问题求解是由一定的情景引起的，按照一定的目标，应用各种认知活动、技能等，经过一系列的思维操作，使问题得以解决的过程。需要作出新的过程的问题求解称为创造性问题求解，而采用现存的过程的问题求解称为常规问题求解。

### 10.5.1　问题空间

思维过程总是体现在一定的活动中，这种活动主要表现为问题求解。因此，通过问题求解过程的分析，研究思维过程。整个解决问题过程中事实上包括记忆、学习、技能，以及情绪、动机等许多环节，是一个复杂的心理活动，即“高级心理过程”。关于这种高级心理过程能否用简单的“机械学习”“刺激-反应”等规律进行解释，一直存在着争论。人在生活中解决复杂而又困难的问题，人认识和适应外界复杂的环境，像这样复杂的活动能否用实验室得出的基本知识加以解释，仍是一个问题。

一个算术题包括一些数字和运算符。运算符包括加、减、乘、除等。在算题中如果答案正确就是合法运算，否则就是错误操作。例如，“(3＋5)×2＝16”就是合法运算。合法的运算要求运算的问题、数目字、运算结果、运算过程都符合求解问题范围的规则。此外，数字本身也有一些特性：奇数、偶数；数字之间的关系有“＞”(大于)、“＜”(小于)、“＝”(等于)等。因此，在提出一个问题时就要确定任务的领域。考试时，主试者可用几种方式向被试者交代任务的领域和范围。主试者首先要交代问题中具体的数目字、运算符；被试者则必须从过去经验中提取奇数、偶数，判断哪些数大，哪些数小，它们之间的关系等。被试者根据所提取的信息，以及数字之间的关系进行推理，产生了自己的问题空间。因此一个课题对主试者来说是“任务领域”；对被试者来说则成为“问题空间”了。

被试者从主试者那里获得信息，知道了课题是什么，运算符是什么，然后，在自己头脑中转换为“问题空间”。在人工智能中，问题求解基本的形式描述如下：

由两个集命 $S$、$A$，和 $S\times A$ 到 $S$ 的部分函数 $f$，以及 $S$ 的二个元 $s_i$ 和 $s_g$ 构成的五元组 $<S,A,f,s_i,s_g>$ 称作问题，$s_g=f(f(\cdots\ (f(f(s_i,a_1),a_2)a_3)\cdots),a_n)$ 那样执行 $A$ 的元 $a_1,\cdots,a_n$，称为那个问题的问题求解。

上述定义中，$S$、$A$、$f$、$s_i$、$s_g$ 分别称为状态空间、操作集合、状态转移函数、初始状态、目标状态。所谓状态是为了描述事物特征的一组变元 $q_0,q_1,\cdots,q_n$ 构成的一个有序组 $<q_0,q_1,\cdots,q_n>$，其中 $q_i$ 的取值不同，即反映了事物的差异或变化。这里的每一个 $q_i$ 都称为分量。这里 $n$ 的有限性并不是一个必须的限制，也可以是无限个分量。引起状态中的某些分量发生变化，从而使问题从一个状态变化到另一个状态的作用称为操作。

这样，一个问题用全部可能的状态及其相互关系的图来描述，称为状态空间表示法。利用状态空间表示法时，问题的求解过程也就转化为在状态空间中寻求从初始状态 $s_i$ 到目标状态 $s_g$ 的路径问题，或者说从 $s_i$ 到 $s_g$ 的操作序列的问题，即

$$a=a_1,a_2,\cdots,a_n$$

$$s_g=f(f(\cdots(f(f(s_1,a_1),a_2)a_3)\cdots),a_n)$$

下面以水壶问题为例，说明状态空间表示的方法。

给定两个水壶,一个可装4加仑水,一个能装3加仑水,水壶上没有任何度量标记。有一水泵可用来往壶中装水。问题是怎样在能装4加仑的水壶里恰好只装2加仑水。

此问题的状态空间可描述为一组整数偶对$(x,y)$,其中$x=0,1,2,3$或4;$y=0,1,2$,或3。$x$表示在4加仑水壶中的含水量,$y$表示3加仑水壶中的含水量。开始状态为(0,0)。目标状态为$(2,n)$,其中$n$为任意值。用来求解该问题的操作可用如下10条规则来描述。

(1) $(X,Y|X<4)\to(4,Y)$把4加仑水壶装满。

(2) $(X,Y|Y<3)\to(X,3)$把3加仑水壶装满。

(3) $(X,Y|X>0)\to(X-D,Y)$从4加仑水壶里倒出一些水。

(4) $(X,Y|Y>0)\to(X,Y-D)$从3加仑水壶里倒出一些水。

(5) $(X,Y|X>0)\to(0,Y)$把4加仑水壶中的水全部倒在地上。

(6) $(X,Y|Y>0)\to(X,0)$把3加仑水壶中的水全部倒在地上。

(7) $(X,Y|X+Y\geqslant 4\land Y>0)\to(4,Y-(4-X))$把3加仑水壶中的水往4加仑水壶里倒,直到4加仑水壶装满为止。

(8) $(X,Y|X+Y\geqslant 3\land X>0)\to(X-(3-Y),3)$把4加仑水壶中的水往3加仑水壶里倒,直至3加仑水壶装满为止。

(9) $(X,Y|X+Y\leqslant 4\land Y>0)\to(X+Y,0)$把3加仑水壶申的水全部倒进4加仑水壶里。

(10) $(X,Y|X+Y\leqslant 3\land X>0)\to(0,X+Y)$把4加仑水壶中的水全部倒进3加仑水壶里。

这里,把操作表示为规则,左部匹配当前状态,右部是应用规则后的新状态。注意,为完整描述这些操作,需弄清在问题陈述中提到的某些假定。如能用水泵装水进壶,能把壶中水倒在地上,能把一水壶里的水倒进另一水壶里,无别的度量器可用等。在把一典型问题陈述从自然语言形式转换成适于程序使用的形式表示时,还要求诸如此类的一些附加的假定。

为求解水壶问题,除上面给出的问题描述外,还需要附带简单的循环控制结构,选择其左部匹配当前状态的某条规则;并按照该规则右部的行为对此状态作适当改变;检查改变后的状态是否为某一目标状态,若不是,则继续该循环。

有许多操作序列能求解水壶问题,如表10.2所示。

**表10.2 求解水壶问题**

| 4加仑水壶中含水的加仑数 | 3加仑水壶中含水的加仑数 | 应用的规则 |
|---|---|---|
| 0 | 0 | |
| 0 | 3 | 2 |
| 3 | 0 | 9 |
| 3 | 3 | 2 |
| 4 | 2 | 7 |
| 0 | 2 | 5 |
| 2 | 0 | 9 |

## 10.5.2 产生式系统

信息加工系统能进行数学运算是以产生式系统为基础的。所谓产生式系统,是指计算机所能执行的一组活动,它的基本原理是一个条件能产生一个动作,即C-A(Condition-Act)。过去在心理学中有一个S—R公式,即刺激-反应公式。产生式系统是从"条件"到

“活动”，其中的活动就相当于反应。每当某个条件出现时就产生某个活动。S—R 与 C—A 有联系，但也有不同。二者共同的地方是当刺激出现时就会出现反应，当条件出现时就会出现“活动”。但是，在“条件”和“活动”的联系上，“条件”并不等于外来的刺激。“条件”有已有的信息，即在短时记忆中已保持了的信息，这是二者的区别。

图 10.6 给出了一个竖式的加法运算。首先，总的目标是做加法。在具体做的时候，是先读 6，将 6 保存在短时记忆中；再读 4；6 和 4 相加得 10，又将 10 保存在短时记忆中。得到 10 之后，就没有必要再保持 6 和 4 了，只保存 10 就够了。这样，竖式加法的过程是：6＋4 =10，10＋7＝17，17＋5＝22，22＋8＝30。每一步脑子里只保持 1 个数和读 1 个数，另外，要有指示，即已经加到哪一步了？下一步该加什么了？从这个例子中可以看到，人在解决计算任务时，记忆包括如下 3 个分活动。

(1) 关于总目标的信息——做加法。

(2) 每一步输入和输出的信息——每一步由两个数相加得出结果。

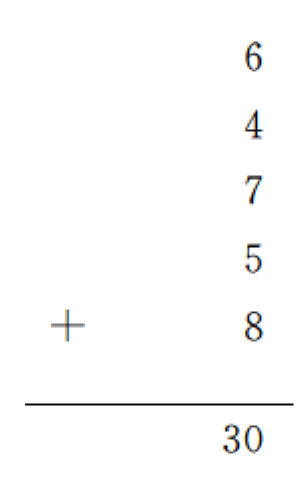



图 10.6　竖式加法

(3) 指示——在解题过程中，指出运算到哪一步。如果已经完成了一步，就把指示指向下一步计算，眼睛由上向下移动。

以上所讲的每一步计算过程就是一个产生式。短时记忆是“条件”，从“条件”得出结果，这个结果又存入短时记忆，并变为下一步的“条件”。在头脑里的并不是具体的“条件”，而是把 10 当成 $N_1$，7 当成 $N_2$，计算 $N_1+N_2$，再往下看后面一个数，等等。这是一般的“条件”和动作之间的关系。儿童学算术，不仅要会算具体数字，而且要学会一般的产生式。信息加工心理学关于产生式系统的基本假设，就是把条件保持在短时记忆里，由条件引起活动；条件在短时记忆里是心理活动；由条件引起的活动也是心理活动。因为 10＋7＝17 的加法实际上是从长时记忆中提取的，相加之后，又将相加的结果 17 放在短时记忆里，为下一步相加时使用。从长时记忆转到短时记忆也是内部活动。

可见所有产生式活动都与短时记忆有关，这包括把信息存入短时记忆，从短时记忆里提取信息，以及信息在短时记忆内部的变化和传递。下面以竖式加法运算为例，列出 C-A 产生式的基本方式(表 10.3)。

**表 10.3　产生式的基本方式**

| 产　生　式 | 计 算 过 程 |
|---|---|
| ① 从短时记忆开始，转到外部运动 | ① 算完一步，眼睛注意下一步 |
| ② 输入感觉信息，并将其放进短时记忆保存 | ② 念一个数字，把它记到短时记忆里 |
| ③ 从短时记忆里提取信息并加以改变，然后又存入短时记忆 | ③ 做完一步，得出暂时的和，再加一步，得出新的和 |
| ④ 从短时记忆里提取信息，将其固定或保持到长时记忆里 | ④ 全部算完以后，把总数记到长时记忆里 |
| ⑤ 从长时记忆里提取信息，放到短时记忆里 | ⑤ 利用早已学会的加法口诀，算出结果 |

目标和指示是产生式条件的控制部分，也是最主要的部分。没有控制部分，计算就无法进行。如果目标改变。进行的方式也发生改变。

在长时记忆里还有一个小部分——程序部分。长时记忆里存有许多数据，它们不仅包括动作，还包括从经验获得的关于世界的知识。因此，一个产生式系统的基本要素如下。

(1) 一组规则，每一规则有左部和右部，左部确定了该规则的可应用性，右部描述了应用该规则时要采取的行动。

(2) 一个或多个数据库，它们含有适于该特定任务的信息。数据库中有些部分可以是永久的，有些部分则仅与求解当前问题有关。可按任何适当的方法构造数据库中的信息。

(3) 一种控制策略，它规定了规则同数据库作比较的次序，还规定了多条规则同时匹配某一数据时应解决的办法。

再举一个例子来说明产生式系统的基本方式。假定我们要解决的是由4个圆盘组成的汉诺塔(Tower of Hanoi)问题。4个圆盘中最大的为1号，依次为2号、3号、4号。最小的4号在最上面(图10.7)。为解决这个问题，被试者要做一系列动作。例如，当被试者要把4号盘移到B柱的3号盘上面，他想到4号盘并将它移到3号盘上面，这是表10.3中第一种从短时记忆提取信息并转到动作的产生式；注意到3号盘在B柱上是表10.3中第二种获得感觉信息并将其转入短时记忆的产生式；事先已看到最小的4号盘在C柱上，并把它作为待移动的目标保持在记忆中，这是从短时记忆到短时记忆的第三种产生式；被试者在短时记忆中将4号盘在C柱上的信息保持约1秒，还要把它转入长时记忆，这是第四种产生式。此外，被试者过去可能解决过这种问题，记得假若把3号盘放到C柱上问题就不能解决了。这是把过去的经验(长时记忆)提取出来放入短时记忆的第五种产生式。由此可知，表10.3中列出的五种动作是最基本的信息加工活动。

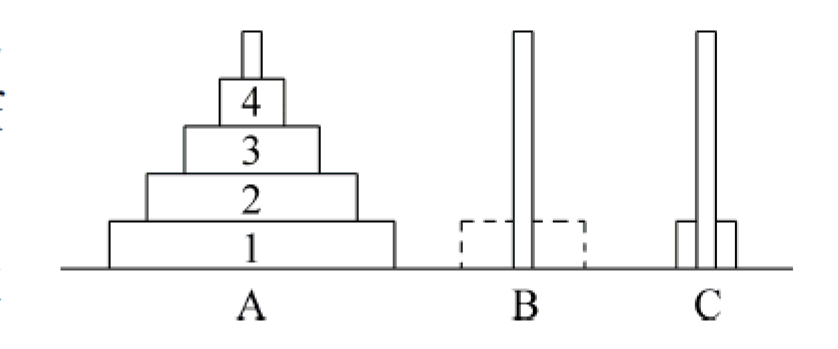


图10.7　汉诺塔问题

产生式系统可用来模仿任一可计算过程。这一点早由波斯特(Post)于1943年证明。1967年明斯基讨论过把波斯特过程用作一可计算模式的问题。1973年安德森在人类联想存储一文中提出“扩散激活理论”。安德森认为，在长时记忆里并没有独立的短时记忆的小盒子，而只有一个小的活动着的部分。例如使用计算机编程时，在长时记忆里过去的有关部分被激活了。被激活了的这一小部分就相当于我们所说的短时记忆。在长时记忆中被激活的部分会不断地扩散、变化，这就相当于从长时记忆中提取短时记忆。

## 10.5.3　启发式搜索

为有效求解难题，通常需要平衡灵活性和系统性这两个不同的要求来构造一种控制策略。这种策略不保证找到最佳解，但常能找到较好的解，人们经常采用启发式搜索，它以牺牲完备性的代价，改进搜索过程的有效性。

对于一个现实的智能系统A来说，它的存在环境B可以按照智能系统A本身具有的知识，划分为如下三大类。

(1) 完全信息环境B。它可以看成是一类问题的集合。系统A已经具备了全部信息，或问题的全部状态空间，它们将任何的初始状态和相应的目标状态联系起来。我们可以设计出一个算法来求解这些问题，这个算法叫A的全信息算子或E型算子，记作Te。全信息

环境 Be 中的问题是确定性问题。

(2) 不完全信息环境，或者部分信息环境 Bp。它也可以看成是一类问题的集合。系统 A 只掌握了把其中的任何初始状态与相应的目标状态联系起来的部分信息。我们只能设计出一个过程来试探求解这类问题，这个过程叫系统 A 的部分信息算子或 P 型算子，记作 Tp。Bp 中的问题是不确定问题，一个不确定问题可以分解成若干子问题时，其中可以包含若干确定的子问题。绝大多数启发式程序都是 P 型算子，但也有少部分是 E 型算子。

(3) 未知环境 Bn。系统 A 对其中的问题一无所知，故 Bn 中的问题叫未知问题。

智能系统所面临的实际问题，大多数是属于部分信息环境中的不确定问题，系统不知道与问题有关的全部信息，因而无法知道该问题的全部状态空间，不可能用一套算法来求解其中的所有问题。只有依靠部分状态空间和一些特殊的经验性规则来求解其中的部分问题。有的问题虽然是全信息环境，但是由于算法效率太低，根本无法实现。为了提高解题效率，必须利用一些经验性的启发式规则。有些启发式法对大量问题领域都通用，有些启发式法则仅代表与一特定问题的解有关的某种专门知识。

启发式信息的表示方式一般有如下两种。

(1) 规则。例如，下棋系统的规则不仅能简单描述一组合法棋步，而且能描述一组由书写规则的人确认的“高明”棋步。

(2) 启发式函数。该启发式函数能对单个问题状态作出估计，以确定其合乎要求的程度。

启发式函数是一映射函数，它把问题状态描述映射成希望的程度，而这种程度通常用数值来表示。考虑问题状态的哪些方面、怎样对所考虑的方面作出估计、给单方面的权如何选择等，都取决于搜索过程中一给定结点处的启发式函数值，对该结点是否在通向问题求解结果的预想路径上作出尽可能好的估计。

启发式函数设计得好，对有效引导搜索过程获得解具有重要作用。有时，非常简单的启发式函数对一路径好与不好能作出相当令人满意的估计。但在有些场合下，则要用非常复杂的启发式函数。下面列出了几个问题的启发式函数。

(1) 下棋：我方超过对方的棋子优势。

(2) 九宫重排：已归位的将牌个数。

(3) 巡回售货员问题：至今的距离总和。

(4) 一字棋：下述值之和。我方能赢的每一行赋值 1，己方有一棋的每一行算值 1，己方有两棋的每一行算值 2。

注意，有时高的启发式函数值表明一个位置相当好(如下棋：九宫重排、一字棋)，而有时低的启发式函数值表明一个有利情况(如巡回售货员问题)。一般来说，以何种方法指出函数关系不大。使用启发式函数值的程序会酌情使之极小或极大。

启发式函数旨在当有多条路径可定时建议该走哪条路径，从而指导搜索过程朝最有利的方向搜索。启发式函数对搜索树(或图)中每一结点的真正优点估计得越精确，解题过程就越少有弯路。在极端情况下，启发式函数能好得让系统不作任何搜索就直接走向某一解。可是对许多问题，计算这种函数值的耗费，会超过由减少搜索过程而节省下来的耗费。但毕竟还是能够通过从所考虑的结点出发做一个完整搜索，并确定它是否通向一个好解的办法，算出一个完美的启发式函数。一般来说，求启发式函数值的耗费与用该函数节省下来的搜

索时间这两者总是有得有失的。

### 10.5.4 手段目的分析法

当我们求解问题时,有的只要采用正向推理搜索策略,有的选择逆向推理搜索策略。然而,常常混合这两种策略,这一混合策略使得能首先求解一个问题的主要部分,然后回过头来求解在把大片断“粘”起来时出现的小问题。手段目的分析技术可以达到这个目的。

手段目的分析法就是先有一个目标,它与人当前的状态之间存在着差异,人认识到这个差异,就要想出某种活动来减小这个差异。但是要完成这个活动,还要先满足某些条件,也就是说要设法减小这方面的差异。手段目的分析法中的“目的”就是“目标”,所谓“手段”就是用什么活动去达到这个目标。

开发手段目的分析法的第一个人工智能程序是通用问题求解程序 GPS。手段目的分析依赖于将一个问题状态转换成另一个问题状态的一组规则;不过,不用按照完整的状态描述来表达这些规则。只需将每条规则表达成左部和右部。左部描述应用此规则应该满足的条件,称为前件。右部描述此规则的应用会改变该问题状态的哪些方面,称为结果。下面列出一个简单的家用机器人所用的操作符。

```
操作符
PUSH(obj,loc)   前件
                  at(robot,obj) ∧ large(obj) ∧ clear
                  (obj) ∧ armempt
                 结果
                  at(obj,loc) ∧ at(robot,loc)
CARRY(obj,loc)   前件
                  at(robot,obj) ∧ small(obj)
                 结果
                  at(obj,loc) ∧ at(robot,loc)
WALK(loc)       前件
                   无
                 结果
                   at(robot,loc)
PICKUP(obj)     前件
                  at(robot,obj)
                 结果
                   holding(obj)
PUTDOWN(obj)    前件
                   holding(obj)
                 结果
                    ¬ 11olding(obj)
PLACE(objl,obj2)  前件
                    at(robot,obj2) ∧ holding(obil)
                   结果
                    on(objl, obj2)
```

表 10.4 是描述每一操作符何时适合的差别表。注意,有时可能有多个操作符能缩小一个给定差别,有时一个给定操作符能缩小多个差别。

假设要该领域里的机器人将写字台从一间房搬到另一间房，放在写字台上的两件东西也随之搬走。则开始状态与目标状态之间的主要差别是写字台的位置。为缩小这一差别，应挑选 Push 或 Carry。

**表 10.4 差别表**

| | Push | Carry | Walk | Pick up | Put down | Place |
|---|---|---|---|---|---|---|
| Move object | √ | √ | | | | |
| Move robot | | | √ | | | |
| Clear object | | | | √ | | |
| Get object on object | | | | | | √ |
| Get arm empty | | | | | √ | √ |
| Be holding object | | | | | | |

若先选 Carry，则要满足其前件。这导出两个要缩小的差别：机器人的位置和写字台的大小。机器人的位置可用 Walk 来处理，但没有操作符能改变物体的大小（因未给操作符 Saw-apart）。因此，该路径通向死胡同。沿另一分支，试应用 Push。图 10.8 就是问题求解程序在此刻的进展。它已找到做某项有用事件的方法。但还不在做那件事项的位置，且做此事尚未完全到达目标状态。因此，现要缩小 A 与 B 之间的差别以及 C 与 D 之间的差别。

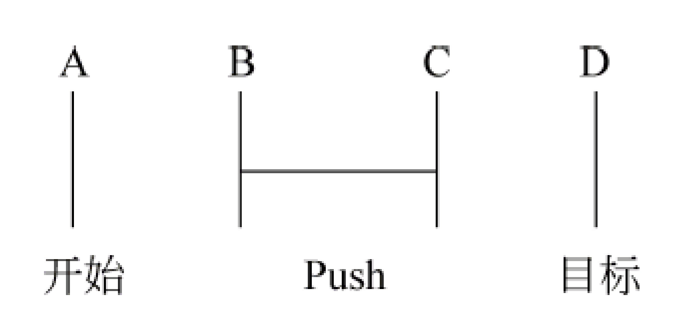



图 10.8 手段目的分析法的进展

Push 有三条前件，其中两条产生开始状态与目标状态之间的差别。因为写字台很大，所以有一前件不建立差别。用 Walk 把机器人带到正确的位置。用两次 Pick up 可清掉写字台表面。但在一次 Pick up 之后，试做第二次时有另一差别，即臂得空。Put down 可用来缩减那一差别。

一旦做完 Push，问题状态就接近但不完全到达目标状态。得把物体放回在写字台上。Place 能做此事，但不能立即应用。另一差别得消去，因机器人要抓物体。图 10.9 就是问题求解程序在此处的进展。

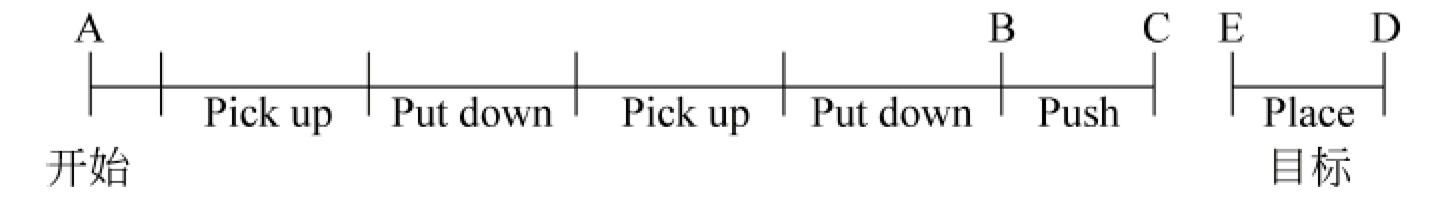



图 10.9 手段目的法的更大进展

先用操作符 Walk 将机器人带回到物体，再用 Pick up 和 Carry，从而缩小 C 与 E 之间的差别。采用“手段目的”分析法解决问题，可以缩小目标和当前状态之间的差异，减小步子，缩短距离，使问题容易得到解决。

## 10.5.5 解决问题的策略

人类的思维主要是为了解决问题。为了提高人类思维的效率，即提高人类产生新知识的效率和扩大人类对现有知识的利用，人们需要研究解决问题的策略。根据启发式信息的来源，解决问题的策略大致可分为两大类。一类是内部指导的策略，即在短时记忆中已经有

了一些信息,以这些信息为指导而达到目标。这就是说被试者用内部已有的信息,而不是用外部的信息。另一种策略叫做刺激指导的策略,或刺激引出的策略。这种策略是被试用者感觉器官感知到当前的情景或刺激条件,然后引出下一步应该怎样做。前一种情况,主要靠内部短时记忆已经存储的信息来指导解决问题的行为顺序和行为模式。例如机械记忆策略、模式策略、目标递归策略都属于这一类。一种情况是根据外界刺激及其引出的反应来解决问题,即一个刺激引起一个反应,再来一个刺激又引起另一个反应。例如知觉策略就是这类解决问题的方法。

### 1. 机械记忆策略

这种方法是把做对的一系列步骤死记硬背下来,以此来解决相同的或者相似的问题。人们通过机械学习获取知识。环境所提供的信息记忆在大脑里,当在下次使用时仅是检索,而不需要重复计算、推理或查询。机械记忆策略是通过接收曾解过的问题,记住问题和它的解,从而进行工作。抽象地看,执行动作可以设想为某个函数 $f$,输入模式是$(X_1,\cdots,X_n)$,计算的输出值是$(Y_1,\cdots,Y_p)$。$f$ 的机械记忆简单地以取想对$[(X_1,\cdots,X_n),(Y_1,\cdots,Y_p)]$记亿在大脑中。后来 $f(X_1,\cdots,X_n)$输入时,只要简单地从大脑中检索$(Y_1,\cdots,Y_p)$,而无须再进行计算。机械学习的简单模型如图 10.10 所示。

$$(X_1,\cdots,X_n) \xrightarrow{f} (Y_1,\cdots,Y_p) \xrightarrow{\text{记忆}} [(X_1,\cdots,X_n),(Y_1,\cdots,Y_p)]$$

输入模式　执行函数　计算输出值　联想对

图 10.10　机械学习的简单模型

假定桌子上有 5 本书,为了叙述方便,我们用 A、B、C 分别表示左、中、右三个位置。这 5 本书从小到大的顺序为 1、2、3、4、5(见图 10.11)。任务是将这 5 本书从 A 挪到 C。如果我们是 A→C→B→A…的顺序移动最小 1 本书,那么,第一步我们把最小的 1 本书(编号为 1)由 A 移到 C,第二步则只能把 2 由 A 移到 B,第三步将 1 由 C 到 B,而第四步则只能将 3 由 A 移到 C。此后,是将 1 由 B 移到 A,将 2 由 B 移到 C…。照这样下去,我们就能把 5 本书都合法地移到 C。这个移动过程的进行顺序可以编成一个押韵的简谱:

1213|1214|1213|1215|1213|1214|1213|121

如果移动的是四本书,只要取上面模式中 121312141213121 这部分即可。在总数为六本书的情况下,只要在前面讲的模式后面加上 6,然后再把 6 前面的模式重复一遍即可。

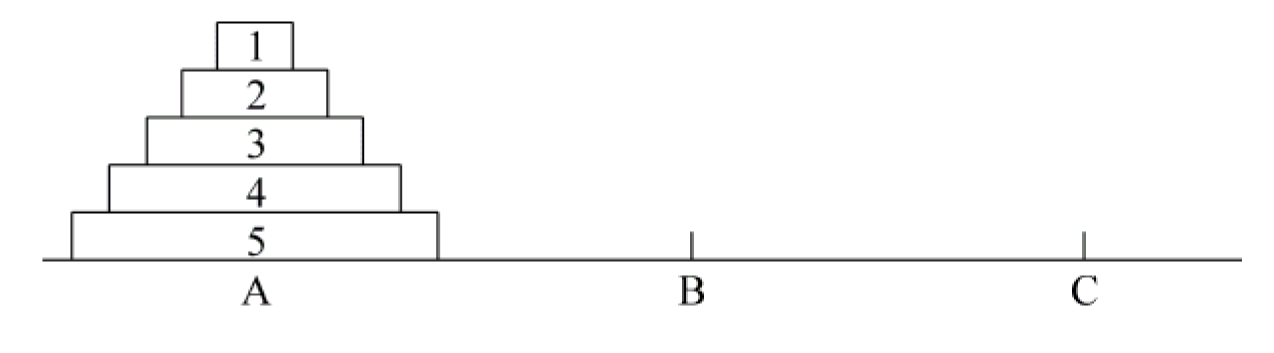



图 10.11　移书的汉诺塔问题

### 2. 模式策略

模式策略在解决问题的过程中没有目标问题,只有几种可能的动作。

(1) 最小的 1 本书总可以向任何方向移动。

(2) 可把没有被最小的书挡住的次小的书移到另一个地方。

因此，为了合法的解决问题，我们就要采用与可能的动作相适应的规则。这些规则是：移动最小一本书以后，下一步不要把这本书挪回到原来的位置。即我们可以按 A→B、B→C、C→A、A→B…的顺序，或按 A→C、C→B、B→A、A→C…的顺序来移动最小的一本书。另一条规则是：在第奇数次移动时，总是移动最小的那本书；而在偶数次移动时则总到移动除最小的那本书以上的书。但是，实际上这时能够合法移动的书只有一本，合法的动作也只有一种。这样就可得出移动过程的顺序：

1213|1214|1213|1215|1213|1214|1213|121

只要按照上面的模式，不用记住目标和步骤就可解决任何数目的汉诺塔问题。

### 3. 目标递归策略

为了解决移动 5 本书的问题，我们先解决移动四本书的问题；为解决移动 4 本书的问题，我们先解决移动 3 本书的问题；为解决移动 3 本书的问题，我们先解决移动 2 本书的问题；为解决移动 2 本书的问题，先解决移动 1 本书的问题。这样计算机程序可采用循环方式。若采用 Prolog 语言实现汉诺塔的问题求解，程序如下所示：

```
• hanoi(N): —move(N,left; centre,right).
   move(o,—,—,—): —!.
   move(N,A,B,C): —
       M is N-1,
       move(M,A,C,B),
       inform(A,B),
   ;   move(M,C,B,A).
   inform(X,Y): —write([move,a,book,from,the,
                    X,pole,to,the,Y,pole]),n1.
```

当书为 5 本时，则 $N$ 等于 5。我们定义谓词 hanoi，它具有一个参数 $N$。这样的 hanoi($N$) 表示当左面的书本为 $N$ 时，打印出移动的顺序。两个 move 子句中第一个，表示边界条件，第二个子句实现递归，循环执行。谓词 move 有四个参数。第一个参数是要移书本的总数。另外三个参数表示原来位置、中间位置、目标位置。谓词 inform 使用 write 打出位置名字，这些位置是书本移动时的情况。

这种策略是内部指导的目标递归策略。我们不必用眼睛看具体的东西，只是把内部目标记在脑子里，就可以一步步循环，最后解决问题。这种策略使短时记忆发挥重要作用，必须记住最终的目的是什么，下一步应该移动什么，现在走到哪一步了等等。

### 4. 知觉策略

知觉策略依赖于外界刺激，是刺激指向的策略。我们的目的是将最大的一本书移到目标位置上，而我们看到最大的一本书还没有在目标位置上。在进行这种策略的操作时，脑子里不必记住目标在哪里，下一步应该移动什么。因为我们可用眼睛看到这一切，视觉告诉我们情景与目标的关系。在这个策略里，我们的目标始终是把下一本没有在 C 上的最大的书移到 C：

```
GOAL: NEXT LARGEST TO C
```

假如所处的情境是：1、2、4 这三本书在位置 B 上，3、5 这两本书在目标 C 上(图 10.12)。现在的目标是把 4 移到位置 C 上。按照规定，由于两个原因我们不能直接把 4 移到 C：

(1) 4 上面压着两本书，我们一次只能移动一本书。

(2) 目标位置 C 处，5 的上面有比 4 小的 3，我们不能把较大的 4 压在较小的 3 上面。

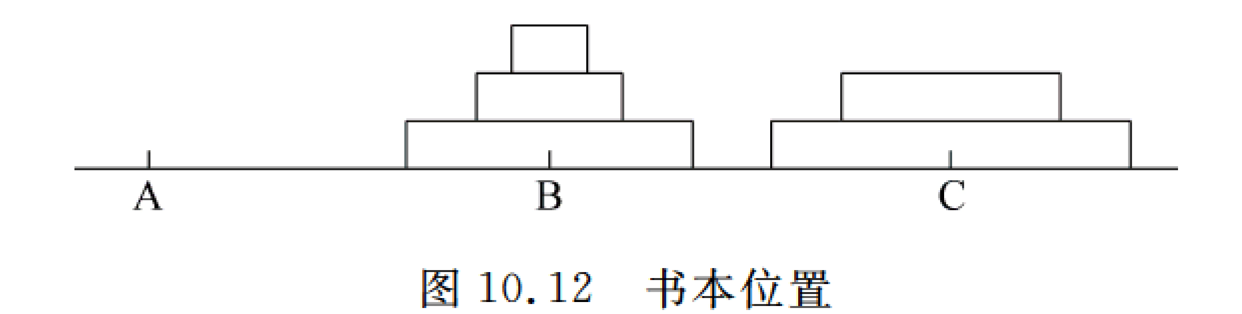



图 10.12　书本位置

我们建立一个目标——移开最大障碍：

GOAL: REMOVE LARGEST BLOCKADER

因为最大的障碍是第三本书，因此应先移走 3。假如 1 和 2 在 5 的上面，3 在 4 的上面，根据目标，还是先移走 3，因为 3 是起阻碍作用的书中最大的 1 本。当把 3 从 C 移到 A 以后，还是不能直接把 4 移到 C，因为 1 和 2 在 4 的上面，这时 2 是最大的障碍，所以现在的目标是移走 2。但是 1 在 2 上面，不能马上移走 2，这时 1 又成了最大的障碍。于是，我们的新目标是移走 1。实际的做法是 M(3,C,A)；M(1,B,C)；M(2,B,A)；M(1,C,A)；M(3,B,C)。在操作过程中，如果有别的事情吸引了我们的注意。那么，在注意完那件事后，回来注意移书的任务。因为我们的眼睛看见目标，随时都知道目标在哪里。另外，在操作过程中，必须明确解决课题的规则，否则操作时会发生混乱。

上述四种策略，都可以解决汉诺塔问题。但是它们在学习时间、对记忆的要求、回忆、转移等方面是不同的。第一种方法的学习时间最长，要求长时记忆的信息多。回忆是学会一种方法后，经过一段时间再把它复述出来。后三种方法因为有一定的图式和模式，比第一种方法容易记住。转移是指用已学到的方法解决类似的新问题的能力。第一种方法是不容易转移的。第二和第三种方法可以解决书本数更多或更少的移动问题。第四种方法稍加修改，也容易转移到新的情境中。因此我们必须择优选择解决问题的策略，以达到事半功倍的效果。

## 10.6　决策理论

决策过程是一个信息流动和再生的过程：在决策的各个阶段，信息在信息源(通过信息载体)和决策者之间交互，将知识、数据、方法等传递给决策者，影响决策的制定；同时，决策形成过程中产生的新知识、新数据、新方法又回流到信息源，经过信息载体的整理加工生成新的信息记录下来，并同时完成信息载体中错误、陈旧信息的修改更新工作；信息对决策的影响还体现在决策实施过程中，信息流可以随时把出现的情况和问题反馈给信息载体，经过信息再生过程后记录下来，用以指导新的决策工作。

管理学家西蒙认为，科学的决策过程至少包括以下四个步骤：找出存在问题，确定决策目标；拟定各种可行的备择方案；分析比较各备择方案，从中选出最合适的方案；决策的执行。信息的高效流动是科学决策的前提条件。图 10.13 表示的是决策过程中的信息流动过程。

信息源是指信息的出处。常见的信息源包括各种类型的出版物、档案资料、会议记录、传媒工具以及重要人物的讲话等。在计算机技术飞速发展的信息时代，各种类型的计算机情报检索数据库的建立，使得远距离快速获取信息成为可能。信息载体包括人脑、语言、文献资料和实物等。信息附着在信息载体上，并通过信息载体发挥作用。

在决策的各个阶段，信息在信息源(通过信息载体)和决策者之间交互，将知识、数据、方法等传递给决策者，影响决策的制定；同时，决策形成过程中产生的新知识、新数据、新方法又回流到信息源，经过信息载体的整理加工生成新鲜的信息记录下来，并同时完成信息载体中错误的、陈旧的信息的修改更新工作。信息对决策的影响还体现在决策实施过程中，信息流可以随时把出现的情况和问题反馈给信息载体，经过信息再生过程后记录下来，用以指导新的决策工作。

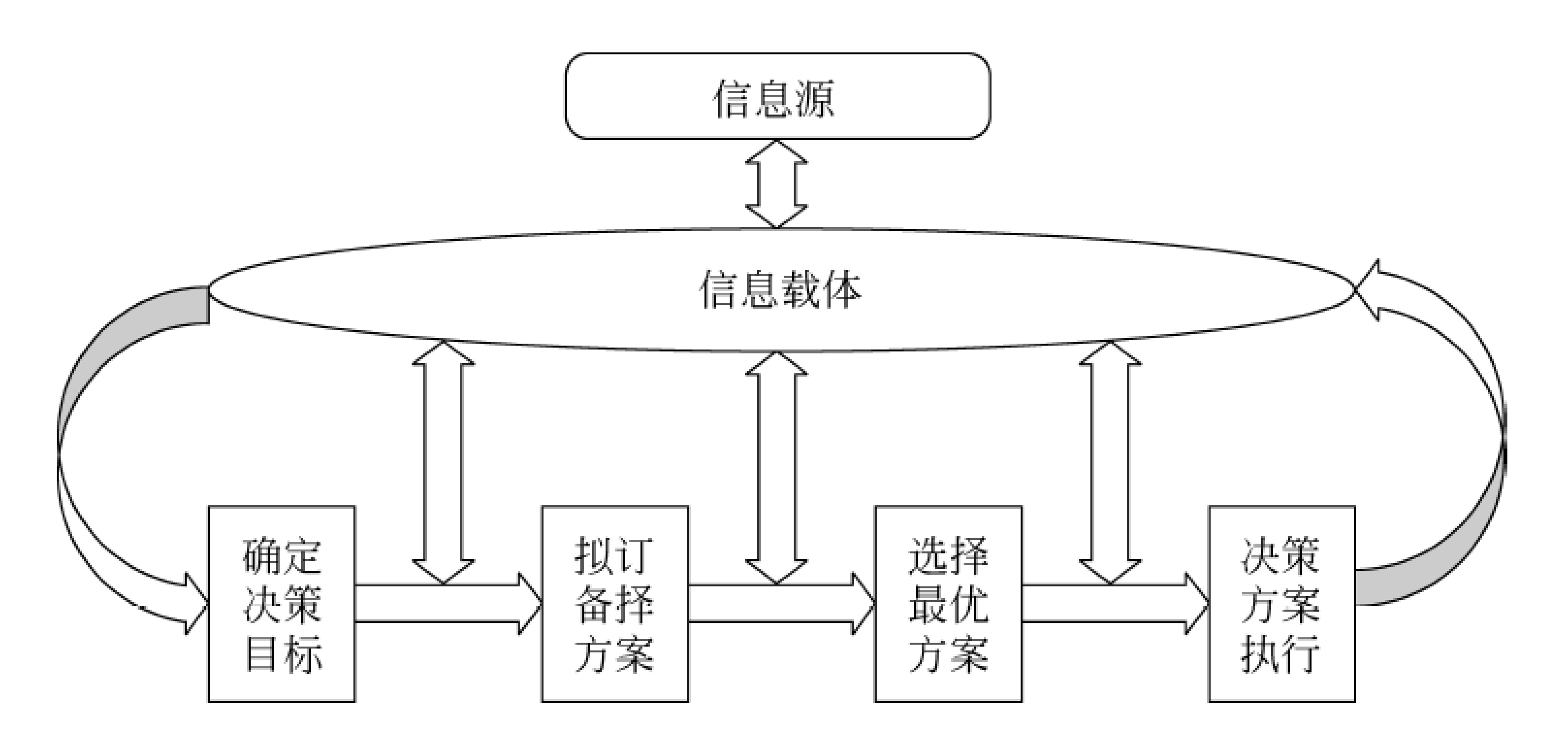



图 10.13 决策过程中的信息流动

信息流动的最终目的是要方便人们做出科学的决策以解决实际问题。信息是决策的基础，但并不是说只要有了信息，就一定可以做出正确的决策，关键在于如何对信息进行科学的加工处理。实际上，整个信息的流动过程也就是一个信息处理和再生的过程。只有在对充分的信息进行适当处理的基础上，才能产生出新的、用以指导行动的策略信息。

## 10.6.1 决策效用理论

关于决策最早的理论称为"经典决策理论"，它们反映了经济学的观点。这些理论假设决策者：

(1) 知晓所有可能的选择，以及每项选择可能带来的后果。

(2) 对各选项之间的细微差异无限敏感。

(3) 在确定选择哪个选项时完全是理性的。

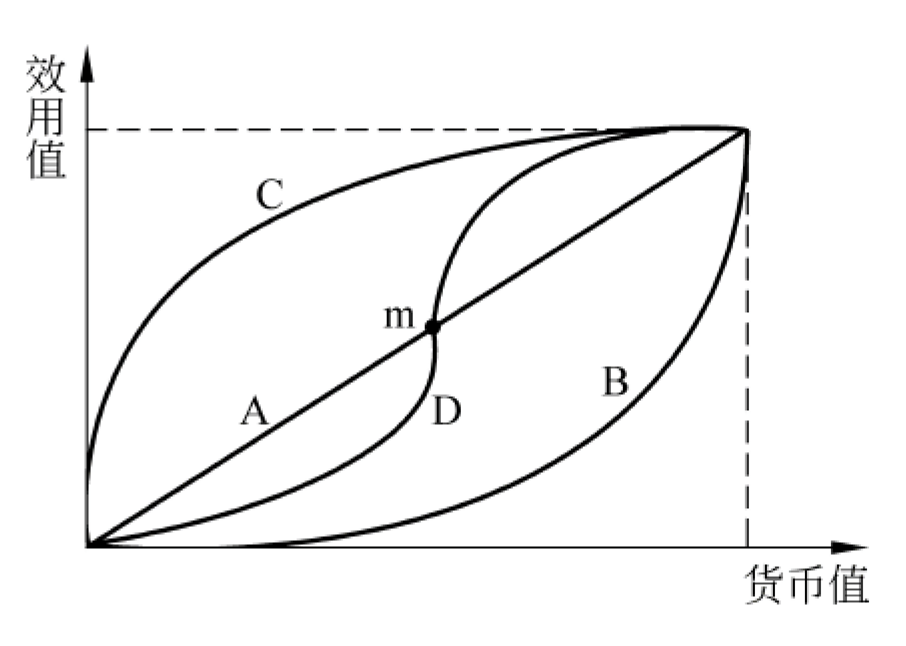



图 10.14 典型的效用函数曲线

决策效用理论主要考虑每个决策者的心理学成分，测出各个决策结果的效用值，并按效用期望值的大小来评价、选择方案。在进行一次性决策(或重复性不大)的风险决策时，应该求出各决策结果的效用值，而效用值可通过效用函数给出。典型的效用函数曲线如图10.14 所示。

曲线 A(中间型)：该效用值与货币值呈线性关

系。具有这种效用函数的决策者,对决策风险持中立态度,或者决策者认为该项决策可以重复进行,因此不必对决策的不利后果特别关注而谨慎从事。由于该效用函数是线性的,效用函数曲线与期望货币曲线重合。

曲线 B(稳妥型):这是减速递增型效用函数,即虽随着货币额的增多效用也递增,但递增的速度越来越慢。具有这种效用函数的决策者对于亏损特别敏感,而大量的收益对他的吸引力却不大,即宁可不赚大钱也不愿意承担大风险。

曲线 C(冒险型):这是加速增加型效用函数,即随着货币额的增多效用也随着递增,而且递增的速度越来越快。具有这种效用函数的决策者十分关注收益而不太顾及风险,敢于冒险,乐于作孤注一掷的大胆尝试。

曲线 D(组合型):这是存在拐点的效用函数。具有这种效用曲线的决策者在货币量不大时具有一定的冒险胆略,但货币量增至一定数量时,决策者就转为采用稳妥策略了。曲线上的拐点 m 就是这一变化的分界线。

人们做出各种合理的决定是根据:

(1) 对所有可能知晓选项都考虑到了,这里假设某些选项是未能预见到的。

(2) 最大程度地利用了已知信息,假设某些相关信息或许未能了解到。

(3) 如果是主观的,仔细地衡量每一选项的潜在代价(风险)和利益。

(4) 仔细地计算各种结果发生的概率,假设结果的必然性不可知。

(5) 在考虑到上述所有因素的基础上最大程度地进行合理推理。

### 10.6.2 满意原则

诺贝尔经济学奖获得者西蒙指出,不是说人类必定无理性,而是有限理性。西蒙提出满意原则的决策策略。在满意原则中,我们并不需要考虑所有可能的选项,也不需要并仔细计算整个选项库中哪一个选项可以最大限度地实现我们的目标,同时使损失最小。相反,我们只是一个接一个地考虑各个选项,一旦我们发现有一个选项可以令我们满意,或者它已经足够好,可以达到我们能够接受的最低水平,此时便立即做出选择。因此,我们只是考虑了最少数量的备择项目便可以做出一个决定,它足以使我们相信它能满足我们的最低要求。

### 10.6.3 逐步消元法

在 20 世纪 70 年代,特沃斯基(Amos Tversky)在西蒙有限理性思想的基础上观察到,当我们面临的选项远远多于我们感觉自己能够合理应对的选项数目时,我们有时会采用另外一种策略。在这种情况下,我们并不会试图对所有可能选项的各个属性都予以考虑。相反,我们会采用逐步消元法:首先集中关注这些选项的某一方面(属性),并且在这个方面制定一个最低标准。对于那些不符合这一标准的选项,我们便可以排除它们了。在剩下的选项,我们继续选择第二个方面并制定其最低标准,以此再去掉一些选项。依此类推,通过从一系列的角度继续使用逐步消元法,直到最后只有一个选项剩下[773]。

特沃斯基观察到,我们经常是在非最优策略的基础上做出决定。他和助手一起经过再三研究最终发现,我们在做决定的时候经常采用某些心理捷径甚至可能是偏见。这些捷径和偏见会限制甚至有时会扭曲我们做出理性决定的能力。我们利用心理捷径的一条关键途

径是以我们的概率估计为中心。

### 10.6.4 贝叶斯决策方法

利用贝叶斯定理求得后验概率，据以进行决策的方法，称为贝叶斯决策方法。贝叶斯定理是关于随机事件 $A$ 和 $B$ 的条件概率和先验概率的概率判断。

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{10.14}$$

式(10.14)中，$P(A|B)$是在 $B$ 发生的情况下 $A$ 发生的可能性。$P(A)$是 $A$ 的先验概率，是根据历史资料或主观判断，未经实验证实所确定的概率，不考虑任何 $B$ 方面的因素。$P(B|A)$是已知 $A$ 发生后 $B$ 的条件概率，也由于得自 $A$ 的取值而被称作 $B$ 的后验概率。$P(B)$是 $B$ 的先验概率，也称作标准化常量。$P(A|B)$是已知 $B$ 发生后 $A$ 的条件概率，由于得自 $B$ 的取值而被称作 $A$ 的后验概率。按这些术语，Bayes 定理可表述为：

后验概率＝(相似度×先验概率)/标准化常量

也就是说，后验概率与先验概率和相似度的乘积成正比。另外，比例 $P(B|A)/P(B)$也有时被称作标准相似度，Bayes 定理可表述为：

后验概率＝标准相似度×先验概率

## 10.7 智能决策支持系统

### 10.7.1 智能决策支持系统

决策支持系统 DSS(Decision Support System)是运筹学、管理科学和计算机科学等结合的产物。智能决策支持系统 IDSS(Intelligent Decision Support System)是人工智能技术和决策支持系统的集成，应用专家系统技术，使 DSS 能够更充分地应用人类的知识，如关于决策问题的描述性知识，决策过程中的过程性知识，求解问题的推理性知识，通过逻辑推理来帮助解决复杂的决策问题的辅助决策系统。智能决策支持系统具有定量分析和定性分析相结合支持决策的能力，能更有效地解决半结构化问题和非结构化问题[914]。

智能决策系统开发平台 IDSDP 是中科院计算所智能科学实验室开发的一种功能强、实用性好的软件工具，为建造大规模管理和决策信息系统提供良好的开发环境。该系统具有下列特点。

(1) 该系统是一个基于多主体的开放系统，每个主体具有自治能力和协同工作方式，系统的可扩展性、可伸缩性好。

(2) 采用客户/服务器(Client/Server)的逻辑连接方式和网络技术，资源共享程度高，性能价格比好。

(3) 该系统将数据库、模型库、知识库、案例库等集成于一体，支持定性与定量分析相结合的科学决策。

(4) 在符号计算和神经计算研究成果的基础上，提供多种分析、预测、决策方法，将多种智能技术融合一体，智能化程度高。

(5) 利用面向对象技术，研制了一种模型描述语言，为用户建立模型提供了一种有用的

工具。系统能根据用户需要自动选择模型。

(6) 提供自然语言、图形、远程访问等人机界面,具有友好的人机交互环境,实现人机共存的决策模式。

智能决策系统开发平台 IDSDP 可以分成五个层次,即通信层、信息源层、管理层、决策层、应用层(见图 10.15)。通信层是支持开放分布处理的重要环节。我们要在 TCP/IP 通信协议的基础上,建立客户/服务器(Client/Server)的系统结构,以支持分布式决策。信息源层包括多种形式的信息资源,即数据、模型、知识、案例等。数据是信息系统的基础,IDSDP 系统中的数据库是基于 Access、SQLServer2000、Oracle 等关系数据库基础上完成的;主要研究交互式数据分析和快速的信息检索。模型是支持决策的核心,模型库是 DSS、IDSS 区别于其他信息系统的重要特征,是进行定量分析所必需的;以研究知识信息处理为对象的知识工程使决策支持系统进入智能决策的新阶段,把领域专家处理问题的经验和知识通过知识获取,建成知识库,运用各种推理策略进行问题求解,使系统可以像专家那样处理问题,实现定性和定量分析相结合的科学决策。所以,知识是智能决策的关键;利用已有的经验,处理当前的问题是一种成功的方法本系统的范例库将为用户提供尽可能多的有用信息。管理层主要实现对各种信息资源的有效管理。在数据库系统中主要研究了交互式数据分析和快速的信息检索,在模型库系统中采用面向对象技术,研制了模型描述语言,用户可以方便地建立所需要的模型。决策层包括分析、预测、决策和报告生成。应用层的重点是提供友好的人机交互环境,研究和开发受限自然语言接口、图形系统、多媒体系统等,使人和机器在决策时协调工作。

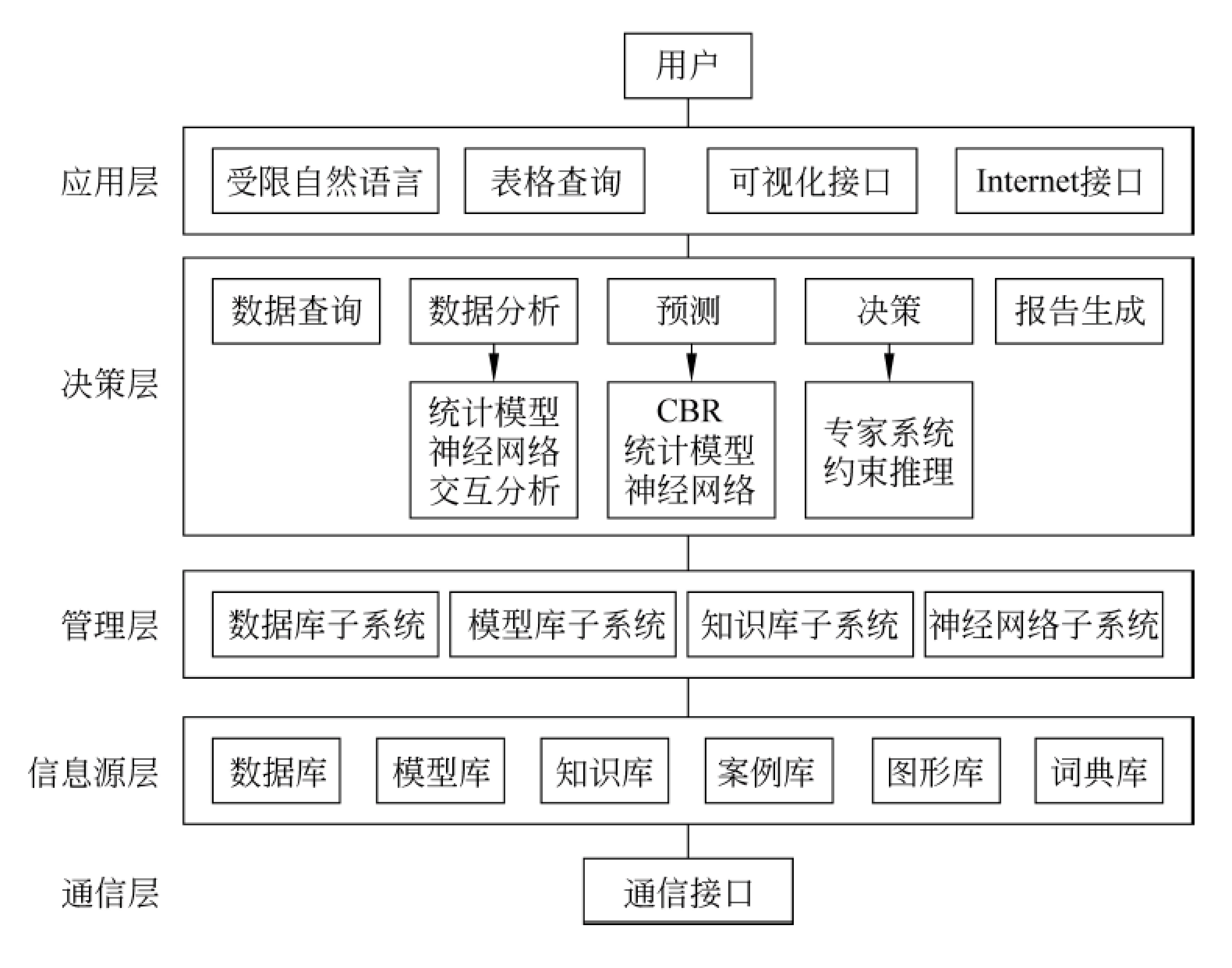



图 10.15 IDSDP 逻辑结构

分布式智能群决策支持系统是一种主体网格智能平台 AGrIP(Agent Grid Intelligent Platform),如图 10.16 所示[696]。主体网格智能平台自低向上可以分为信息资源层、管理层、决策层和应用层。

信息资源层包括所有的数据资源,如 Web 信息、流媒体信息、空间/地理信息等,以及各

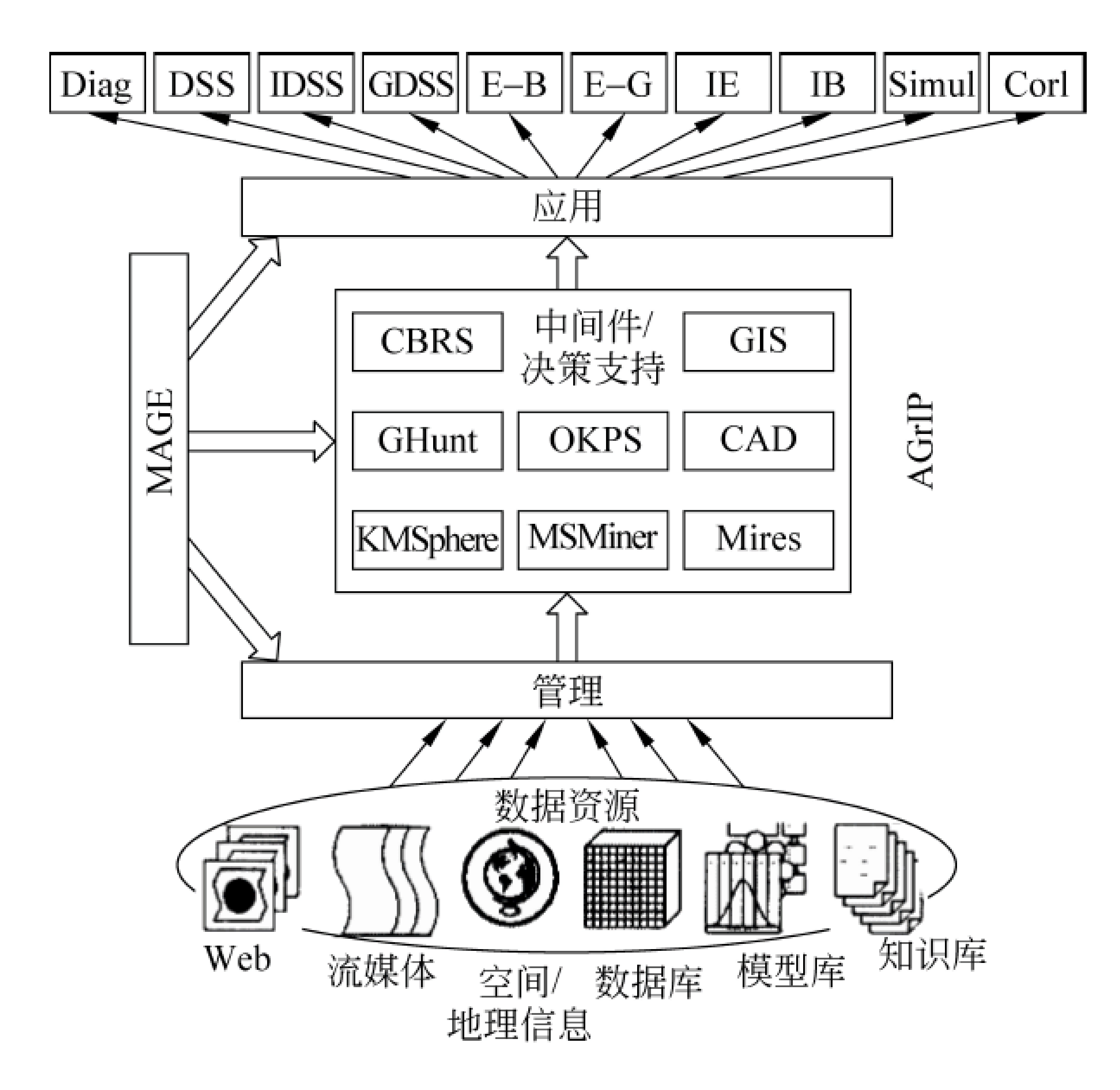



图 10.16 分布式智能群决策支持系统

类数据库、模型库、知识库。数据是信息系统的基础，本系统支持交互式数据分析和快速的信息检索；模型是支持决策的核心，是进行定量分析所必需的；以研究知识信息处理为对象的知识工程使决策支持系统进入智能决策的新阶段，把领域专家处理问题的经验和知识通过知识获取，建成知识库，运用各种推理策略进行问题求解，使系统可以像专家那样处理问题，实现定性和定量分析相结合的科学决策。

管理层主要实现对各种信息资源的有效管理。在数据库系统中主要研究了交互式数据分析和快速的信息检索；在模型库系统中采用面向对象技术，研制了模型描述语言，用户可以方便地建立所需要的模型；在知识库系统中提出了面向对象的知识表示方法，以及基于约束的推理机制。

决策层提供各种支持决策的中间件，具有自主版权的面向对象的专家系统工具 OKPS、数据挖掘平台 MSMiner、基于范例的推理系统 CBRS、多媒体信息检索系统 Mires、智能搜索引擎 GHunt，以及知识管理平台 KMSphere，也可以兼容常用的地理信息系统 GIS 和计算机辅助设计系统 CAD 等系统。

应用层的重点是提供友好的人机交互环境。有了中间件的强有力支持，基于分布式智能群决策支持系统的二次开发就可以只关注于用户的决策问题和环境需求，无须考虑具体的执行过程。此外，由于该平台对图形系统、多媒体系统等的支持，使人和机器在决策时可以更好地协调工作。

当前大数据、云计算、深度学习成为研究热点。云计算是运用计算机网络的传输能力，将信息处理过程从独立的计算机或服务器转移到云端，将网络连接的资源形成资源池，进行统一管理和调度。云计算的软件即服务模式，为决策支持的应用与推广提供了良好的平台。

## 10.7.2 综合集成研讨厅

20 世纪 80 年代钱学森提出了开放的复杂巨系统（Open Complex Giant Systems，

OCGS)的概念,以及对几类开放的复杂巨系统进行了深入的探讨,提出“从定性到定量的综合集成法”作为处理开放的复杂巨系统的方法论,着眼于人的智能与计算机的高性能两者结合,以思维科学与人工智能为基础,用信息技术和网络技术构建“综合集成研讨厅(Hall for Workshop of Metasynthetic Engineering)”体系[856]。

综合集成研讨厅体系可以视为一个由专家体系、机器体系、知识体系三者共同构成的一个虚拟工作空间,如图 10.17 所示。

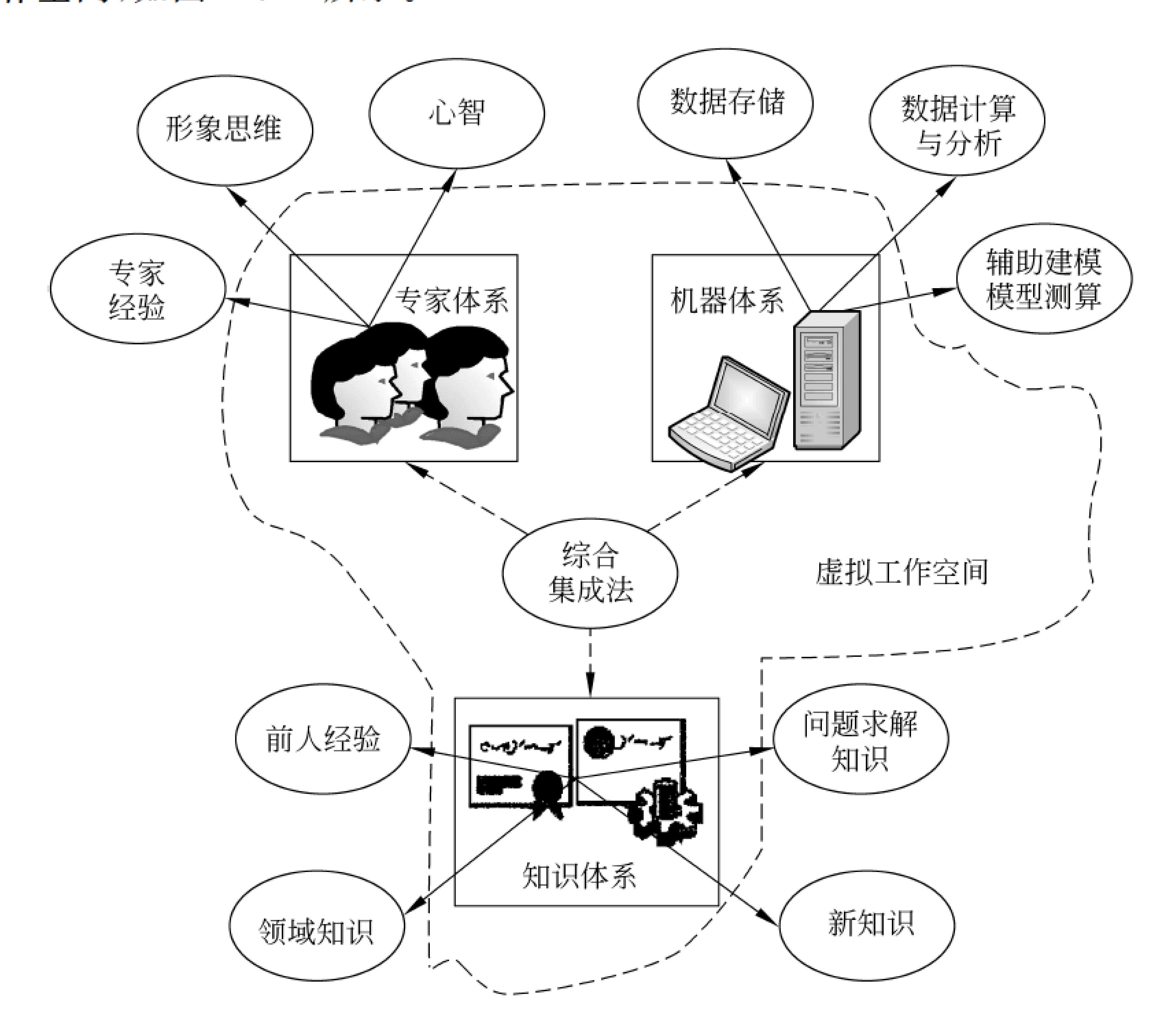



图 10.17　综合集成研讨厅框架结构示意图

在综合集成研讨厅中专家的心智、经验、形象思维能力及由专家群体互相交流、学习而涌现出来的群体智慧在解决复杂问题中起着主导作用;另一方面,机器体系的数据存储、分析、计算以及辅助建模、模型测算等功能是对人心智的一种补充,在问题求解中也起着重要作用。知识体系则可以集成不在场的专家以及前人的经验知识、相关的领域知识、有关问题求解的知识等,还可以是由这些现有知识经过提炼和演化,形成新的知识,使得研讨厅成为知识的生产和服务体系。

综合集成研讨厅是一类巨型智能工程系统。在这个系统中,参加研讨的人与人、人与计算机、计算机与计算机密切合作、借助网络与数据仓库等技术,对所要解决的复杂问题进行研讨与论证。在互联网基础上,通过专家群体的有效互动,使群体的智能涌现出来,从而给出在一定程度上能满足要求的解决方案。

# 第 11 章<<
# 智力发展

智力发展是智能科学研究的重要内容之一。本章首先说明智力的实质、智力的差异、智商以及智力的发展特征，然后介绍不同心理学家如何看待智力，讨论智力能否测量以及如何做到精确和客观的评估。最后还要谈到哪些因素决定了一个人的智力高低，以及智力在多大程度上取决于遗传基因或环境条件等。皮亚杰(J. Piaget)对儿童认知发展领域，如语言、思想、逻辑、推理、概念形成、道德判断等长期的临床研究，创立了以智力发展阶段理论为核心的智力发展理论。

## 11.1 引言

人类认知是一个处于互动过程中的复杂系统，认知系统产生、编码、转换或者处理各种不同类型的信息。在研究认知发展的理论方面，皮亚杰的理论可谓是该领域的出发点。他提出了许多其他理论家迄今还在探究的重要问题。他通过四个阶段描述认知质的变化：感知运动阶段、前运算阶段、具体运算阶段和形式运算阶段。某种动作逻辑变成心理的逻辑运算，这种逻辑运算应用于越来越抽象的表征中。在同化和顺化过程中，系统在大量的环境经历过程中主动地创造了一个关于现实世界的心理结构，而不仅仅是对所经历事物做一个简单的心理复制。与环境世界间的每一次认知冲突常常具有两个方面，即同化与顺化。同化实质上是按照个体已有的认知系统对外在资料的解释或分析；环境际遇经过认知转化，从而与系统已拥有的知识和思考方式相一致。顺化意味着对认知系统作出稍微的改变，以便顾及外在资料的结构。

皮亚杰之后，并且正是由于皮亚杰的影响，认知发展学家沿着五个主要方向开展工作。新的观点探究：

(1) 比阶段更为有限的认知结构的可能性。

(2) 儿童的表现随同一领域内的不同任务而变化，或随不同领域而变化的原因。

(3) 变化过程。

(4) 生物影响和过程。

(5) 认知的社会方面[225]。

新皮亚杰主义融合了皮亚杰的观点、信息加工理论，有时还包括社会文化理论。与皮亚杰不同，他们研究的是认知技能的领域特殊性，心理容量发展的变化，以及对认知活动的社会支持。例如，凯斯(R. Case)强调的是心理容量和问题解决策略在认知发展中的作用。导致认知发展的过程包括对环境中更多元素的加工和协调、对信息的区分，以及将设置次级目

标作为达成最终目标的手段。一组领域特殊性的核心概念结构,负责调集诸如对客体的探索,对他人的观测和模仿,以及在解决诸如人或社会认知等某一特定领域的问题时与他人的合作等活动。

信息加工理论的出发点是信息在一个类似于计算机的系统内的流动。人类注意信息,将其转化为某种信息表征,将其与系统已有的信息加以比较,赋予一定的含义并加以储存。儿童加工速度的提高,以及由此导致的容量的增长,促进了认知的发展。儿童对刺激编码的灵活性和全面性的提高,以及各种新策略的获得,也是重要的变化来源。然而,能够加工多少信息的限制严重束缚了儿童的发展。研究者通过计算机模拟或者对行为进行详细而精确的描述,对认知发展的各种假设进行检验。西格勒的叠波模型反映了儿童思维的可变性和连续的认知变化。信息加工连接主义或神经计算,各种各样强度的连接模式的变化,构成了认知变化。连接主义者强调对大脑的类比。

各种生物理论有望揭示大脑发展的作用。新近在神经成像方面的进展,激发人们对认知神经发展科学产生相当浓厚的兴趣。大脑变化和行为变化之间的相关,暗示着两者之间的双向影响,在某一个发展时刻大脑制约和促进思维,行为反过来决定着大脑接收的刺激的性质。先天模块论采纳的是相当激进的观点,认为某些基本的概念是先天具有的。每个模块是某一个特定领域所特有的,诸如语言、面孔再认或物理客体,并且只是相当松散地与其他模块相联系。

目前,理论观点十分活跃,特别在心理理论领域。研究者研究儿童非正式的直觉的关于世界的"理论",或连贯的因果解释框架。各个概念包含于这些较大的理论中。越后面的理论越复杂。幼儿可能只有一些理论,而年长儿童可能具有各种各样领域的理论。一个理论包含一组关于某一领域内的实体和这些实体之间关系的信念。特别是,理论不同于其他类型的心理表征,因为理论是解释性的,它们可以回答"为什么"的问题。

动力系统框架理论试图把整个认知系统结合起来。因为这种系统是自组织的,所以系统根据自身的当前状态和系统所处的特定环境现场装配某个概念和行动。一个新的行为出现于或"脱落"于这一母体。该理论考查了系统每一时刻的变化,正是这种变化构成了认知变化。

源自维果茨基的社会文化理论,将处于社会情境中的儿童作为分析的主要单元。社会文化历史的综合影响和邻近的社会影响,尤其是父母和其他具有重要意义的成人,是认知变化的主要来源。成人和年长的同伴对儿童解决问题进行指导、支持、鼓励和纠正,从而推动儿童跨越最近发展区。通过使儿童参与指导性的活动,社会帮助儿童达到他们认知功能的最高水平。通过观察具有较高认知能力的他人,以及在成人的指导下尝试新的技能,儿童像学徒那样主动地学习许多内容。

所有这些目前均十分活跃的各种不同理论,为我们提供了更加全面的视野,表明儿童如何通过主动建构类别、规则、认知结构、技能、理论和程序而认识世界。生物和环境既促发又制约这种发展。儿童通过同化、顺化、编码以及与成人或同伴的协同建构,而建构知识。

限于篇幅,本书仅以皮亚杰的理论及其相关概念为主,讨论有关智力发展问题。

# 11.2　智力理论

智力(Intelligence)是什么？迄今为止心理学家尚未能提出一个为众人所接受的明确定义。有人认为，智力主要是抽象思维的能力；也有心理学家将智力解释为“适应能力”“学习能力”“获得知识的能力”“认识活动的综合能力”。更有某些智力测验的先驱者认为：“智力就是智力测验的那个东西。”心理学家对智力所下的定义，大致可分为三类。

(1) 智力是个体适应环境的能力。个体对其所生活的环境，尤其对变化莫测的新环境愈能适应的人，则其智力愈高。

(2) 智力是个体学习的能力。凡个体能对新事物的学习较易、较快，又能利用经验解决困难问题的人，则其智力较高。这种定义，在学校教育上具有实际的意义。

(3) 智力是个体抽象思维的能力。凡个体能由具体事物获得概念，能运用概念作逻辑推理、判断，则表示其智力较高。

当代著名测验学家魏斯勒(Wechsler)综合上面三种意见，将智力定义为：智力是个体有目的的行为，合理的思维，以及有效的适应环境的综合能力。

以上各种定义，尽管有的强调某一侧面，有的重视全体，但有两个方面是共同的。

(1) 智力是一种能力，而且是属于潜在的能力。

(2) 这种能力通过行为表现。表现方式，或者是适应环境、学习、抽象思维等行为的单独表现，或由此三种行为的综合表现。换言之，智力可被看作个体对事、物、情景各方面表现的功能，而此种功能是由行为而表现。

## 11.2.1　智力的因素论

### 1. 智力的二因论

英国心理学家斯皮尔曼(C. Spearman)在 20 世纪初最早对智力问题进行了探讨。他发现，几乎所有心理能力测验之间都存在正相关。斯皮尔曼提出，在各种心理任务上的普遍相关是由一个非常一般性的心理能力因素或称 g 因素所决定。在一切心理任务上，都包括一般因素(g 因素)和某个特殊因素(或称 s 因素)两种因素。g 因素是人的一切智力活动的共同基础，s 因素只与特定的智力活动有关。一个人在各种测验结果上所表现出来的正相关，是由于它们含有共同的 g 因素；而它们之间又不完全相同，则是由于每个测验包含着不同的 s 因素。斯皮尔曼认为，g 因素就是智力，它不能直接由任何一个单一的测验题目度量，但可以由许多不同测验题目的平均成绩进行近似的估计。

### 2. 流体智力和晶体智力说

20 世纪中期以后，卡特尔(Raymond Cattell)提出了流体智力和晶体智力理论。他认为，一般智力或 g 因素可以进一步分成流体智力和晶体智力两种。流体智力指一般的学习和行为能力，由速度、能量、快速适应新环境的测验度量，如逻辑推理测验、记忆广度测验、解决抽象问题和信息加工速度测验等。晶体智力指已获得的知识和技能，由词汇、社会推理以及问题解决等测验度量。

卡特尔认为,流体智力的主要作用是学习新知识和解决新异问题,它主要受人的生物学因素影响;晶体智力测量的是知识经验,是人们学会的东西,它的主要作用是处理熟悉的、已加工过的问题。晶体智力一部分是由教育和经验决定的,一部分是早期流体智力发展的结果。

到 20 世纪 80 年代,进一步的研究发现,随着年龄的增长,流体智力和晶体智力经历不同的发展历程。和其他生物学方面的能力一样,流体智力随生理成长曲线的变化而变化,在 20 岁左右达到顶峰,在成年期保持一段时间以后,开始逐渐下降;而晶体智力的发展在成年期不仅不下降,反而在以后的过程中还会有所增长。由于流体智力影响晶体智力,它们彼此相关,因此,我们可以假想,不管人的能力有多少种,也不论要处理的任务性质如何,在一切测验分数或成绩的背后,存在一种类似于 g 因素的一般心理能力。在大多数智力测验中,均包括偏重于测量晶体智力和流体智力的两类题目。

### 3. 智力多因素论

美国心理学家瑟斯顿(L. L. Thurstone)于 1938 年对芝加哥大学的学生实施了 56 个能力测验,他发现,某些能力测验之间具有较高的相关,而与其他测验的相关较低,它们可归为 7 个不同的测验群:字词流畅性、语词理解、空间能力、知觉速度、计数能力、归纳推理能力和记忆能力。瑟斯顿认为,斯皮尔曼的二因素理论不能很好地解释这种结果,而且过分强调 g 因素也达不到区分个体差异的目的。因此,他提出智力由以上 7 种基本心理能力构成,并且各基本能力之间彼此独立,这是一种多因素论。根据这种思想,瑟斯顿编制了基本心理能力测验。研究结果发现,7 种基本能力之间都有不同程度的正相关,似乎仍可以抽象出更高级的心理因素,也就是 g 因素。

## 11.2.2 多元智力理论

多元智力理论是由美国心理学家加德纳(Gardner)提出的。他认为,智力的内涵是多元的,由 7 种相对独立的智力成分所构成。每种智力成分都是一个单独的功能系统,这些系统可以相互作用,产生外显的智力行为。这 7 种智力如下。

(1) 言语智力,渗透在所有语言能力之中,包括阅读、写文章以及日常会话能力。

(2) 逻辑——数学智力,包括数学运算与逻辑思维能力,如做数学证明题及逻辑推理。

(3) 空间智力,包括导航、认识环境、辨别方向的能力,比如查阅地图和绘画等。

(4) 音乐智力,包括对声音的辨别与韵律表达的能力,比如拉小提琴或作曲等。

(5) 身体运动智力,包括支配肢体完成精密作业的能力,比如打篮球和跳舞等。

(6) 人际智力,包括与人交往且能和睦相处的能力,比如理解别人的行为、动机或情绪。

(7) 内省智力,对自身内部世界的状态和能力具有较高的敏感水平,包括认识自己并选择自己生活方向的能力。

## 11.2.3 智力结构论

美国心理学家吉尔福特(J. P. Guilford)认为,智力活动可以区分出三个维度,即内容、操作和产物,这三个维度的各个成分可以组成为一个三维结构模型。智力活动的内容包括

听觉、视觉(所听到、看到的具体材料,如大小、形状、位置、颜色)、符号(字母、数字及其他符号)、语义(语言的意义概念)和行为(本人及别人的行为)。它们是智力活动的对象或材料。智力操作指智力活动的过程,它是由上述种种对象引起的,包括认知(理解、再认)、记忆(保持)、发散思维(寻找各种答案或思想)、聚合思维(寻找最好、最适当、最普通的答案)和评价(做出某种决定)。智力活动的产物是指运用上述智力操作所得到的结果。这些结果可以按单元计算(单元),可以分类处理(分类),也可以表现为关系、转换、系统和应用。由于三个维度的存在,人的智力可以在理论上区分为 5×5×6=150 种(见图 11.1)。

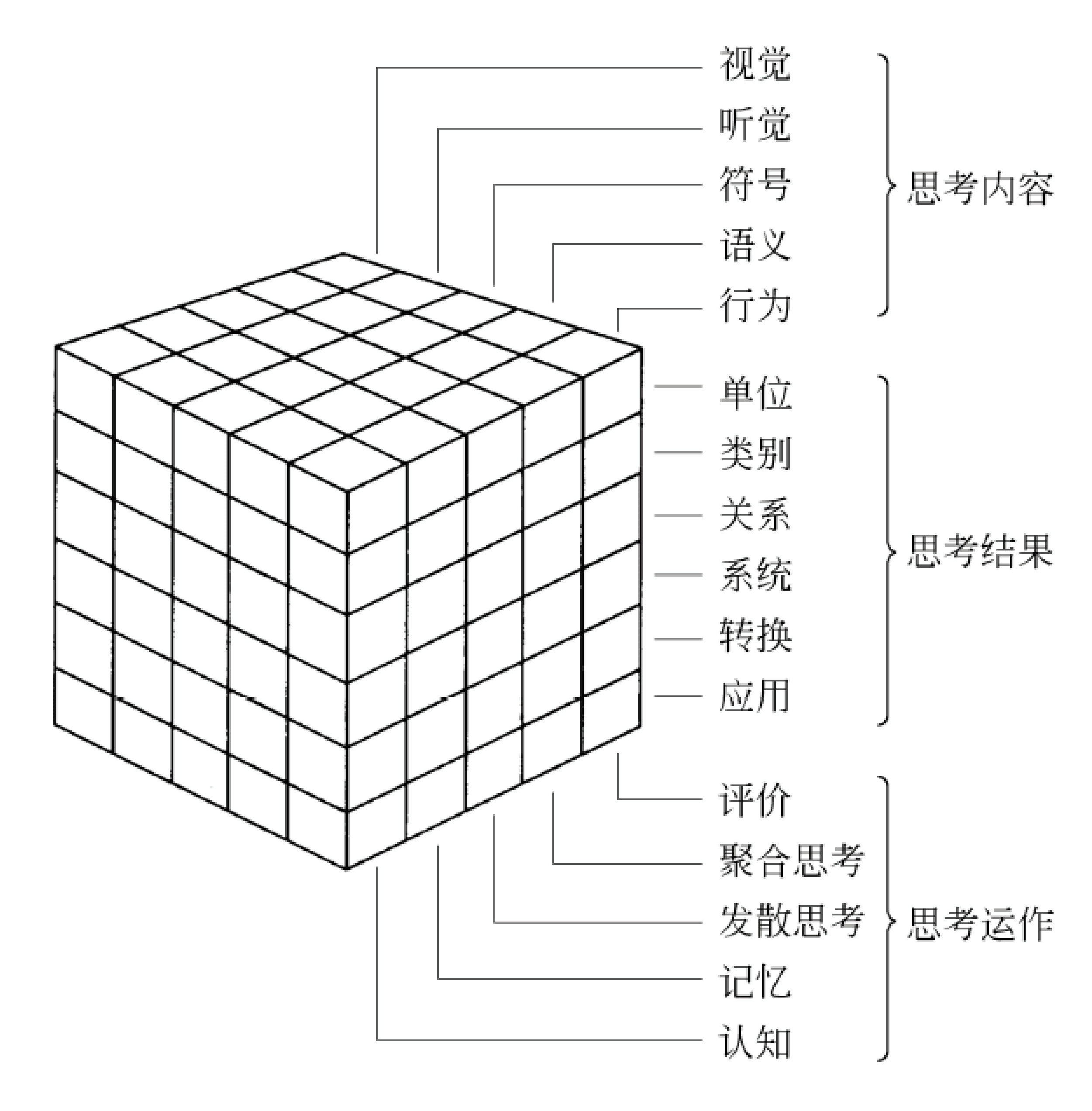



图 11.1　三维智力结构模型

吉尔福特的三维智力结构模型同时考虑到智力活动的内容、过程和产物,这对推动智力测验工作起了重要的作用。1971 年吉尔福特宣布,经过测验已经证明了三维智力模型中的近百种能力。这一成就对智力测验的理论与实践,无疑是巨大的鼓舞。

## 11.3　智力的测量

从智力测验的观点看,这种行为表现智力的观念是极为重要的。因此,有些心理学家干脆把智力定义为:智力乃是一种智力测验的对象。如果进一步追问:智力测验所测量的对象是什么?这个问题虽不易回答,但有一点是肯定的,即所测的对象绝非智力本身,而仍是个体表现在外的行为。间接的测量个体表现在外的行为特征,并量化了,以推估其智力的高低,这是智力测量的基本原则。智力本身只是一个抽象的概念,无法直接测量,这正如物理学上的"能",必须经由物体运动所做的功予以衡量,是同样的道理。

在 20 世纪初,法国心理学家比奈(A. Binet)受巴黎教育当局的委托,承担编制一套测验用来鉴定智力缺陷的学生,让他们能够进入不教授标准课程的学校。从那以后智力测验就首先用来帮助预测儿童和学生的能力,预测他们在"智力"训练中获益多少。现在越来越倾

向于编制和应用智力测验去测定人的能力的不同方面。对于智力测验的主要要求就是要把人有益于准确地按照能力的类别进行分组。这也有赖于智力理论的研究和新智力测验的创造。

智力测验有许多不同的种类。如按被试人数分成个人测验和团体测验;以限定时间内做出正确反应的数目决定分数的为速度测验,而以成功地完成的作业的难度决定分数的为才能测验;要求以言语回答问题的言语测验,非言语的动作反应的作业测验。不管哪种类型的智力测验,一般都包括数量较大、内容不同的测验项目或作业。智力测验的分数就根据成功地完成作业的数目来确定。

智力测验的每一项目都能提供与之适合的年龄水平值。当测试一个儿童时,他所得到的分数是以他通过的项目的数目为依据的。因此他的分数可以用年龄来表示。例如,特曼·梅里尔的要求给每一个词下定义的测验作业,有60%的13岁儿童能做对,就将这一项目在测验中规定给13岁。

假设一个孩子通过了10岁儿童测试过的测验的全部项目,他还通过了11岁的和12岁的某些项目。首先以他10岁以前的所有项目(10岁的包括在内)记分。11岁的项目则通过一半,12岁的项目通过了四分之一,那么,他的记分就再加上六个月(11岁的)和三个月(12岁的)。把他的分数全部加在一起为10年9个月,这就是智龄(MA)。因此,智龄是根据智力测验的作业成绩换算而得的,它是由所通过的测验项目的难度水平决定的。

智商(IQ)的定义是智龄除以实足年龄,然后乘以100,公式如下:

$$\text{智商(IQ)} = \frac{\text{智龄(MA)}}{\text{实足年龄(CA)}} \times 100 \tag{11.1}$$

公式中乘以100的目的一方面可以消去小数,使所得的智商成为整数;另一方面可显示出智力的高低。这种确定智商的方法是假定智力年龄是同实际年龄一起增长的。如果相反,在到达某一实际年龄时智力年龄不再增长,那么一个人若已到此年龄,此后他的年龄再增长时,则他所得的智商就越来越小。但实际上他的智力并未减少。人到达一定的实际年龄之后,智力年龄的发展就停留在相对稳定的水平。由于在15岁时智力不再与实际年龄成正比增加,因此对15岁及15岁以上的人所用求智商的公式为:

$$\text{IQ} = \frac{\text{MA}}{15} \times 100 \tag{11.2}$$

这种方法也不能取得满意的结果。魏斯勒(D. Wechsler)提出成人智力量表,主要内容如下:

(1) 性质及内容:在性质上,这个测验项目分为言语和作业两类。前者包括常识、理解、算术、类同、记忆广度、词汇六个分测验,共计84题。后者包括物形配置、填图、图系排列、按照图案搭积木、符号替换五个分测验,共计44题。两者合起来为128题。题目性质广泛,所测者为个人普通能力。

(2) 适用范围:适用于16岁以上的成人。

(3) 实施程序:个别实施,全测验约需1小时。

(4) 记分与标准:各分测验的原始分数经过换算手续变为加权分数。前六个分测验的加权分数之和即为言语量表的总分。后五个分测验的加权分数之和即为作业量表的总分。两个量表的总分相加即为全测验的总分。量表总分可再按年龄组别查对照表而求得标准分数智商。

本测验标准之建立是根据具有代表性的 700 人所组成的标准化样本。在此样本中对性别(男女各半)、年龄(16～64 岁)、地域、种族、职业以及教育程度等因素均有适当分配,故其代表性甚高。

(5) 信度与相关系数:由折半法求得信度系数是:言语量表 0.96,作业量表 0.93,全量表 0.97。相关系数的研究以斯坦福-比奈量表(简称"斯比量表")为标准,求得相关系数为:言语量表 0.83,作业量表 0.93,全量表为 0.85。

个体智力存在差异。如对大量未经选择的人,施以智力测验,得到的智商分布如表 11.1 所示。这是采用 1973 年修订的斯比量表,对 2～18 岁的 2904 人进行智力测验的结果。由智商分布表看出,智商极高及极低的均占少数,大多数人的智力属于中等或接近中等。

**表 11.1　智商分布**

| 智　　商 | 类　　别 | 百分比 |
|---|---|---|
| 140 及以上 | 极优 | 1 |
| 120～139 | 优异 | 11 |
| 110～119 | 中上 | 18 |
| 90～109 | 中等 | 46 |
| 80～89 | 中下 | 15 |
| 70～79 | 临界 | 6 |
| 70 以下 | 弱智 | 3 |

美国心理学家特尔门(L. M. Terman),采用追踪观察的方法来研究智力超常儿童的才能发展[753]。在 1921—1923 年间,特尔门选择 1528 名智商超过 130 的中小学生,其中男生 857 人,女生 671 人。他对所有对象都做了学校调查和家庭访问,详细了解老师和家长对他们智力的评价,还对三分之一的人作了体格检查。1928 年他回访这些学生所在的学校和家庭,了解他们进入青少年时期以后的智力发展和变化情况。1936 年,这些研究对象都已经长大成人,各自走上了不同岗位。特尔门继续采用通信的方式进行随访,掌握他们的才能发展状况。1940 年,他特地把这些研究对象邀集到斯坦福大学来座谈,并且做了一次心理测验。以后,他仍旧坚持每隔 5 年做一次通信调查,直到 1960 年。

特尔门逝世以后,美国心理学家西尔斯等人继续进行这项研究。1960 年,这些研究对象的平均年龄已达 49 岁。西尔斯作了一次通信调查,人数是原先的 80%。1972 年,他再次进行了通信随访,被调查的人数仍旧保持在原先的 67%。这时,他们的平均年龄已经超过 60 岁。

这样研究前后持续了半个世纪,积累了大量的宝贵资料。研究表明:早期智力超常并不能保证成年以后具备杰出的才能,卓有建树;一个人的能力大小同儿童期的智力高低关系不大;有才能、有成就的人并不都是老师和家长认为十分聪明的人,而是那些长年锲而不舍、精益求精的人。由于这项研究成果在心理学上具有重大意义,美国心理学协会在 1976 年把卓越贡献奖授予这项研究。

怎样鉴别优秀儿童和学生呢?美国学者里思提出以下 17 项作为进行鉴别的心理学准则。

(1) 知识和技能：具有基本技巧和知识，能够适当应用这些技巧解决具体问题。

(2) 注意力集中：不容易分心，能在充分的时间里对一个问题集中注意力来求得解决的办法。

(3) 热爱学习：喜欢探讨问题和做作业。

(4) 坚持性：把指定的任务作为重要目标，用急切的心情去努力完成它。

(5) 反应性：容易受到启发，对成人的建议和提问都能作出积极反应。

(6) 理智的好奇心：从自己解答问题中得到满足，并且能够提出新的问题。

(7) 对挑战的反应：乐意处理比较困难的问题、作业和进行争论。

(8) 敏感性：具有超过年龄的机灵性和敏锐的观察力。

(9) 口头表达的熟练程度：善于正确地应用众多的词汇。

(10) 思维流畅：能够形成许多概念，善于适应新的比较深刻的概念。

(11) 思维灵活：能够摆脱自己的偏见，用他人的观点看问题。

(12) 独创性：能够用新颖的或者异常的方法去解决问题。

(13) 想像力：能够独立思考，富有想像力。

(14) 推理能力：能够把给定的概念推广到比较广泛的关系中去，能够从整体的关系中去理解给定的材料。

(15) 兴趣广泛：对各种学问和活动都感兴趣，如艺术、戏剧、书法、阅读、数学、科学、音乐、体育活动和社会常识。

(16) 关心集体：乐于参加各种集体活动，助人为乐，和他人融洽相处，对别人不吹毛求疵。

(17) 情绪稳定：经常保持自信、愉快和安详，有幽默感，能够适应日常变化，不暴怒。

## 11.4 皮亚杰认知发展理论

1882 年德国生理学家和心理学家普莱尔(W. Preyer)的《儿童心理》一书问世，标志了科学儿童心理学的诞生。在这之后一百多年来，各国心理学家们对儿童智力成长过程进行了大量观察和研究，在这些人当中有盖塞尔(A. Gesell，自然成熟论)、弗洛伊德、华生(J. B. Watson，行为主义)以及埃里克森(E. H. Erikson，人格发展渐成说)等，他们的工作增进了人们对儿童智力发展的理解，同时也构成了当今儿童发展心理的主要流派，其影响是巨大而深远的。

皮亚杰的心理学，从实验到理论，都有自己独到之处。皮亚杰学派对儿童的语言、判断、推理、因果观、世界观、道德观念、符号、时间、空间、数、量、几何、概率、守恒、逻辑等问题进行了大量的实验研究，为儿童心理学、认知心理学或思维心理学开辟了新园地，提出了一套全新的学说。对当代儿童心理学产生广泛而又深刻的影响。

根据皮亚杰的推论，人类生来就有组织和适应倾向，人类将事物予以系统地组合使之成为系统严密的整体，称为组织倾向。人类对环境适应或调整称为适应倾向。人类的智能过程将经验转换成适应新情境所需的认知结构，与生物学过程将食物消化并转换成身体所需的能量一样。人类的认知过程力求均衡作用，与生物学过程维持平衡相同。均衡作用是一种自动调节作用，它使人类所获的概念得到稳定。适应倾向是通过调整和同化两种相互配

合的作用。调整是改变自己的认知结构或认知模式，以适应新的经验。同化是融合新的经验于现存的认知结构里。皮亚杰却对孩子是如何犯错误的思维过程进行了长期的探索，皮亚杰发现分析一个儿童对某问题的不正确回答比分析正确回答更具有启发性。采用临床法(Clinical Method)，皮亚杰先是观察自己的三个孩子，之后与其他研究人员一起，对成千上万的儿童进行观察，他找出了不同年龄儿童思维活动质的差异以及影响儿童智力的因素，进而提出了独特的儿童智力阶段性发展理论，引发了一场儿童智力观的革命，虽然这一理论在很多方面目前也存在争论，但正如一些心理学家所指出：这是“迄今被创造出来的唯一完整系统的认知发展理论”。

## 11.4.1　图式

皮亚杰认为智慧是有结构基础的，而图式(Schema)就是他用来描述智慧(认知)结构的一个特别重要的概念[572]。皮亚杰对图式的定义是“一个有组织的、可重复的行为或思维模式”。凡在行动可重复和概括的东西我们称之为图式。简单地说：图式就是动作的结构或组织。图式是认知结构的一个单元，一个人的全部图式组成一个人的认知结构。初生的婴儿，具有吸吮、哭叫及视、听、抓握等行为，这些行为是与生俱来的，是婴儿能够生存的基本条件，这些行为模式或图式是先天性遗传图式，全部遗传图式的综合构成一个初生婴儿的智力结构。遗传图式是图式在人类长期进化的过程中所形成的，以这些先天性遗传图式为基础，儿童随着年龄的增长及机能的成熟，在与环境的相互作用中，通过同化、顺应及平衡化作用(后述)，图式不断得到改造，认知结构不断发展。在儿童智力发展的不同阶段，有着不同的图式。如在感知运动阶段，其图式被称为感知运动图式，当进入思维的运算阶段，就形成了运算思维图式。

图式作为智力的心理结构，是一种生物结构，它以神经系统的生理基础为条件，目前的研究还无法指出这些图式的生理性质和化学性质。相反，这些图式在人的头脑中的存在是可以根据观察到的行为推测的。事实上，皮亚杰是根据大量的，通过临床法所观察到的现象，结合生物学、心理学、哲学等学科的理论，运用逻辑学以及数学概念(如群、群集、格等)来分析描述智力结构的。由于这种智力结构符合逻辑学和认识论原理，因此图式不仅是生物结构，更重要的是一种逻辑结构(主要指运算图式)。尽管诸如前述视觉抓握动作的神经生理基础是新神经通路髓鞘形成，而髓鞘形成似乎是遗传程序的产物。包含着遗传因素的自然成熟也确实在使儿童智慧发展遵循不变的连续阶段的次序方面起着不可缺少的作用，但在从婴儿到成人的图式发展中，成熟并不起决定作用。智慧演变为一种机能性的结构，是诸多因素共同作用的结果，儿童成长过程中智力结构的完整发展不是由遗传程序所决定。遗传因素主要为发展提供了可能性，或是说对结构提供了门径，在这些可能性未被提供之前，结构是不可能演化的。但是在可能性与现实性之间，还必须有一些其他因素，例如练习、经验和社会的相互作用。

还必须指出，皮亚杰所提出的智力结构具有三要素：整体性、转换性和自动调节性。结构的整体性指结构具有内部融贯性，各成分在结构中的安排是有机的联系，而不是独立成分的混合，整体和部分都由一个内在规律所决定。一个图式有一个图式的规律，由全部图式所构成的儿童的智力结构并非各个图式的简单相加。结构的转换性指结构并不是静止的，而是有一些内在的规律控制着结构的发展，儿童的智力结构，在同化、顺应、平衡化作用下，不

断发展,体现了这种转换性。结构的自调性是指结构由于其本身的规律而自行调节,结构内的某一成分的改变必将引起结构内部其他成分的变化。只有作为一个自动调节的转换系统的整体,才可被称为结构。

同化与顺应是皮亚杰用于解释儿童图式的发展或智力发展的两个基本过程。皮亚杰认为"同化就是外界因素整合于一个正在形成或已形成的结构",也就是把环境因素纳入机体已有的图式或结构之中,以加强和丰富主体的动作。也可以说,同化是通过已有的认知结构获得知识(本质上是旧的观点处理新的情况)。例如,学会抓握的婴儿当看见床上的玩具,会反复用抓握的动作去获得玩具。当他独自一个人,玩具又较远婴儿手够不着(看得见)时,他仍然用抓握的动作试图得到玩具,这一动作过程就是同化,婴儿用以前的经验来对待新的情境(远处的玩具)。从以上解释可以看出,同化的概念不仅适用于有机体的生活,也适用于行为。顺应是指"同化性的格式或结构受到它所同化的元素的影响而发生的改变"。也就是改变主体动作以适应客观变化。也可以说,改变认知结构以处理新的信息(本质上即改变旧观点以适应新情况)。例如上面提到那个婴儿为了得到远处的玩具,反复抓握,偶然地,他抓到床单一拉,玩具从远处来到了近处,这一动作过程就是顺应。

皮亚杰以同化和顺应释明了主体认知结构与环境刺激之间的关系,同化时主体把刺激整合于自己的认知结构内,一定的环境刺激只有被个体同化(吸收)于他的认知结构(图式)之中,主体才能对之作出反应。或者说,主体之所以能对刺激作出反应,也就是因为主体已具有使这个刺激被同化(吸收)的结构,使得这个结构具有对之作出反应的能力。认知结构由于受到被同化刺激的影响而发生改变,这就是顺应,不作出这种改变(顺应),同化就无法运行。简言之,刺激输入的过滤或改变叫做同化,而内部结构的改变以适应现实就叫做顺应。同化与顺应之间的平衡过程,就是认识的适应,也是人的智慧行为的实质所在。

同化不能改变或更新图式,顺应则能起到这种作用。但皮亚杰认为,对智力结构的形成主要有功的机能是同化。顺应使结构得到改变,但却是同化过程中主体动作反复重复和概括导致了结构的形成。

运算是皮亚杰理论的主要概念之一。在这里运算指的是心理运算。什么是运算?运算是动作,是内化了的、可逆的、有守恒前提、有逻辑结构的动作。从这个定义中可看出,运算或心理运算有四个重要特征。

(1) 心理运算是一种在心理上进行的,内化了的动作。例如,把热水瓶里的水倒进杯子里去,倘若我们实际进行这一倒水的动作,就可以见到在这一动作中有一系列外显的、直接诉诸感官的特征。然而对于成人和一定年龄的儿童来说,可以用不着实际去做这个动作,而在头脑里想像完成这一动作并预见它的结果。这种心理上的倒水过程,就是所谓"内化的动作",是动作能被称为运算的条件之一。可以看出,运算其实就是一种由外在动作内化而成的思维,或是说在思维指导下的动作。新生婴儿也有动作,哭叫、吸吮、抓握等,这些动作都是一些没有思维的反射动作,所以不能算做运算。事实上,由于运算还有其他一些条件,儿童要到一定的年龄才能出现有称为运算的动作。

(2) 心理运算是一种可逆的内化动作。这里又引出可逆的概念。可以继续用上面倒水过程的例子加以解释,在头脑中我们可以将水从热水瓶倒入杯中,事实上我们也能够在头脑中让水从杯中回到热水瓶去,这就是可逆性(Reversibility),是动作成为运算的又一个条件。一个儿童如果在思维中具有了可逆性,可以认为其智慧动作达到了运算水平。

(3) 运算是有守恒性前提的动作。当一个动作已具备思维的意义,这个动作除了是内化的可逆的动作,它同时还必定具有守恒性前提。所谓守恒性(Conservation)是指认识到数目、长度、面积、体积、重量、质量等尽管以不同的方式或不同的形式呈现,但保持不变。装在大杯中的 100 毫升水倒进小杯中仍是 100 毫升,一个完整的苹果切成 4 小块后其重量并不发生改变。自然界能量守恒、动量守恒、电荷守恒都是具体的例子。当儿童的智力发展到了能认识到守恒性,则儿童的智力达到运算水平。守恒性与可逆性是内在联系着的,是同一过程的两种表现形式。可逆性是指过程的转变方向可以为正或为逆,而守恒性表示过程中量的关系不变。儿童思维如果具备可逆性(或守恒性),则差不多可以说他们的思维也具备守恒性(或可逆性)。否则两者都不具备。

(4) 运算是具有逻辑结构的动作。前面介绍过,智力是有结构基础的,即图式。儿童的智力发展到运算水平,即动作已具备内化、可逆性和守恒性特征时,智力结构演变成运算图式。运算图式或者说运算不是孤立存在的,而是存在于一个有组织的运算系统之中。一个单独的内化动作并非运算而只是一种简单的直觉表象。而事实上动作不是单独的、孤立的,而是互相协调的、有结构的。例如,人们为了达到某种目的而采取动作,这时一般需要动作与目的有机配合,而在达到目的的过程中形成动作结构。在介绍图式时,已说过运算图式是一种逻辑结构,这不仅因为运算的生物学生理基础目前尚不清楚而是由人们推测而来,更重要的是因为这种结构的观点是符合逻辑学和认识论原理的。因为是一种逻辑结构,故心理运算又是具有逻辑结构的动作。

以运算为标志,儿童智力的发展阶段可以分为前运算时期和运算时期;继之又可将前者分为感知运动阶段和表象阶段;后者区分为具体运算阶段和形式运算阶段。

## 11.4.2　儿童智力发展阶段

皮亚杰将儿童从出生后到 15 岁智力的发展划分为四个发展阶段。对于发展的阶段性,皮亚杰概括了三个特点。

(1) 阶段出现的先后顺序固定不变,不能跨越,也不能颠倒。它们经历不变的、恒常的顺序,并且所有的儿童都遵循这样的发展顺序,因而阶段具有普通性。任何一个特定阶段的出现不取决于年龄而取决于智力发展水平。皮亚杰在具体描述阶段时附上了大概的年龄只是为了表示各阶段可能出现的年龄范围。事实上,由于社会文化不同,或文化相同但教育不同,各阶段出现的平均年龄有很大差别。

(2) 每一阶段都有独特的认知结构,这些相对稳定的结构决定儿童行为的一般特点。儿童发展到某一阶段,就能从事水平相当的各种性质的活动。

(3) 认知结构的发展是一个连续构造(建构)的过程,每一个阶段都是前一阶段的延伸,是在新水平上对前面阶段进行改组而形成新系统。每阶段的结构形成一个结构整体,它不是无关特性的并列和混合。前面阶段的结构是后面阶段结构的先决条件,并为后者取代。

### 1. 感知运动阶段(出生至 2 岁左右)

自出生至 2 岁左右,是智力发展的感知运动阶段。在此阶段的初期即新生儿时期,婴儿所能做的只是为数不多的反射性动作。通过与周围环境的感觉运动接触,即通过他加以客体的行动和这些行动所产生的结果来认识世界。也就是说,婴儿仅靠感觉和知觉动作的手

段来适应外部环境。这一阶段的婴儿形成了动作格式的认知结构。皮亚杰将感知运动阶段根据不同特点再分为六个分阶段。从刚出生时婴儿仅有的诸如吸吮、哭叫、视听等反射性动作开始，随着大脑及机体的成熟，在与环境的相互作用中，到此阶段结束时，婴儿渐渐形成了随意有组织的活动。下面简介六个分阶段。

第一分阶段（反射练习期，出生至1个月）：婴儿出生后以先天的无条件反射适应环境，这些无条件反射是遗传决定的，主要有吸吮反射、吞咽反射、握持反射、拥抱反射及哭叫、视听等动作。通过反复地练习，这些先天的反射得到发展和协调。发展与协调意味着同化与顺应的作用。皮亚杰详细观察了婴儿吸吮动作的发展，发现吸吮反射动作的变化和发展。例如母乳喂养的婴儿，如果又同时给予奶瓶喂养，可以发现婴儿吸吮橡皮奶头时的口腔运动截然不同于吸吮母亲乳头的口腔运动。由于吸吮橡皮奶头较省力，婴儿会出现拒绝母乳喂养的现象，或是吸母乳时较为烦躁。在推广母乳喂养过程中应避免给婴儿吸橡皮奶头可能正是这一原因。从中也可以看出婴儿在适应环境中的智力增长：他愿吸省力的奶瓶而不愿吸费力的母乳。

第二分阶段（习惯动作和知觉形成时期，1～4个月）：在先天反射动作的基础上，通过机体的整合作用，婴儿逐渐将个别的动作联结起来，形成一些新的习惯。例如婴儿偶然有了一个新动作，便一再重复。如吸吮手指，手不断抓握与放开，寻找声源，用目光追随运动的物体或人，等等。行为的重复和模式化表明动作正在同化作用中，并开始形成动作的结构，反射运动在向智慧行动过渡。由于行为并没有什么目的，只是由当前直接感性刺激来决定，所以还不能算作智慧行动。但是婴儿在与环境的相互适应过程中，顺应作用也已发生，表现为动作不完全是简单的反射动作。

第三分阶段（有目的动作逐步形成时期，4～9个月）：从4个月开始，婴儿在视觉与抓握动作之间形成了协调，以后儿童经常用手触摸、摆弄周围的物体，这样一来，婴儿的活动便不再限于主体本身，而开始涉及对物体的影响，物体受到影响后又反过来进一步引起主体对它的动作，这样就通过动作与动作结果造成的影响使主体对客体发生了循环联系，最后渐渐使动作（手段）与动作结果（目的）产生分化，出现了为达到某一目的而行使的动作。例如一个多彩的响铃，响铃摇动发出声响引起婴儿目光寻找或追踪。这样的活动重复数次后，婴儿就会主动地用手去抓或是用脚去踢挂在摇篮上的响铃。显然可以看出，婴儿已从偶然地、无目的地摇动玩具过渡到了有目的地反复摇动玩具，智慧动作开始萌芽。但这一阶段目的与手段的分化尚不完全、不明确。

第四分阶段（手段与目的分化协调期，9～11、12个月）：这一时期又称图式之间协调期。婴儿动作目的与手段已经分化，智慧动作出现。一些动作格式（图式）被当作目的，另一些动作格式则被当作手段使用。如儿童拉成人的手，把手移向他自己够不着的玩具方向，或者要成人揭开盖着玩具的布。这表明儿童在做出这些动作之前已有取得物体（玩具）的意向。随着这类动作的增多，儿童运用各动作格式之间的配合更加灵活，并能运用不同的动作格式来对付遇到的新事物，就像以后运用概念来了解事物一样，婴儿用抓、推、敲、打等多种动作来认识事物。表现出对新的环境的适应，儿童的行动开始符合智慧活动的要求。不过这阶段婴儿只会运用同化格式中已有的动作格式，还不会创造或发现新的动作顺应世界。

第五分阶段（感知动作智慧时期，12～18个月）：这一时期的婴儿，皮亚杰发现，能以一种试验的方式发现新方法达到目的。当儿童偶然地发现某一感兴趣的动作结果时，他将不

只是重复以往的动作,而是试图在重复中作出一些改变,通过尝试错误,第一次有目的地通过调节来解决新问题。例如婴儿想得到放在床上枕头上的一个玩具,他伸出手去抓却够不着,想求助爸爸妈妈可又不在身边,他继续用手去抓,偶然地他抓住了枕头,拉枕头过程中带动了玩具,于是婴儿通过偶然地抓拉枕头得到了玩具。以后婴儿再看见放在枕头上的玩具,就会熟练地先拉枕头再取玩具。这是智慧动作的一大进步。但儿童不是自己想出这样的办法,他的发现是来源于偶然的动作中。

第六分阶段(智慧综合时期,18～24 个月):这个时期儿童除了用身体和外部动作来寻找新方法之外,还能开始"想出"新方法,即在头脑中有"内部联合"方式解决新问题,例如把儿童玩的链条放在火柴盒内,如果盒子打开不大,链条能看得见却无法用手拿出,儿童于是便会把盒子翻来覆去看,或用手指伸进缝道去拿,如手指也伸不进去,这时他便会停止动作,眼睛看着盒子,嘴巴一张一合做了好几次这样的动作之后突然他用手拉开盒子口取得了链条。在这个动作中,儿童的一张一合的动作表明儿童在头脑里用内化了的动作模仿火柴盒被拉开的情形,只是他的表象能力还比较差,必须借助外部的动作来表示。这个拉开火柴盒的动作是儿童"想出来的"。当然儿童此前看过父母类似的动作,而正是这种运用表象模仿别人做过的行为来解决眼前的问题,标志着儿童智力已从感知运动阶段发展到了一个新的阶段。

感知运动阶段,儿童智慧的成长突出地表现在三方面。

(1) 逐渐形成物体永久性(不是守恒)的意识,这与婴儿语言及记忆的发展有关,物体永久性具体表现在:当一个物体(如爸爸妈妈、玩具)在他面前时,婴儿知道这个人或物,而当这个物体不在眼前时,他能认识到此物尽管当前摸不着、看不见也听不到,但仍然是存在的。爸爸妈妈离开了,但婴儿相信他们还会出现,被大人藏起的玩具还在什么地方,翻开毯子,打开抽屉,还应该可以找到。这标志着稳定性客体的认知格式已经形成。近年的研究表明,儿童形成母亲永久性的意识较早,并与母婴依恋有关。

(2) 在稳定性客体永久性认知格式建立的同时,儿童的空间-时间组织也达到一定水平。因为儿童在寻找物体时,他必须在空间上定位来找到它。又由于这种定位总是遵循一定的顺序发生的,故儿童又同时建构了时间的连续性。

(3) 出现了因果性认识的萌芽,这与物体永久性意识的建立及空间-时间组织的水平密不可分。儿童最初的因果性认识产生于自己的动作与动作结果的分化,然后扩及客体之间的运动关系。当儿童能运用一系列协调的动作实现某个目的(如拉枕头取玩具)时,就意味着因果性认识已经产生了。

### 2. 前运算阶段(2～7 岁)

与感知运动阶段相比,前运算阶段儿童的智慧在质方面有了新的飞跃。在感知运动阶段,儿童只能对当前感觉到的事物施以实际的动作及思维,在阶段中、晚期,形成物体永久性意识,并有了最早期的内化动作。到前运算阶段,物体永久性的意识巩固了,动作大量内化。随着语言的快速发展及初步完善,儿童频繁地借助表象符号(语言符号与象征符号)来代替外界事物,重视外部活动,儿童开始从具体动作中摆脱出来,凭借象征格式在头脑里进行"表象性思维",故这一阶段又称为表象思维阶段。前运算阶段,儿童动作内化具有重要意义。为说明内化,皮亚杰举过一个例子:有一次皮亚杰带着 3 岁的女儿去探望一个朋友,皮亚杰

的这位朋友家也有一个 1 岁多的小男孩,正放在婴儿围栏(Playben)中独自嬉玩,嬉玩过程中婴儿突然跌倒在地下,紧接着便愤怒而大声地哭叫起来。当时皮亚杰的女儿惊奇地看到这情景,口中喃喃有声。三天后在自己的家中,皮亚杰发现 3 岁的小姑娘似乎照着那 1 岁多小男孩的模样,重复地跌倒了几次,但她没有因跌倒而愤怒啼哭,而是咯咯发笑,以一种愉快的心境亲身体验着她在三天前所见过的"游戏"的乐趣。皮亚杰指出,三天前那个小男孩跌倒的动作显然早已经内化于女儿的头脑中了。

在表象思维的过程中,儿童主要运用符号(包括语言符号和象征符号)的象征功能和替代作用,在头脑中将事物和动作内化。而内化事物和动作并不是把事物和动作简单地全部接受下来而形成一个摄影或副本。事实上内化是把感觉运动所经历的东西在自己大脑中再建构,舍弃无关的细节(如上例皮亚杰的女儿并没有因跌倒而愤怒啼哭),形成表象。内化的动作是思想上的动作而不是具体的躯体动作。内化的产生是儿童智力的重大进步。

皮亚杰将前运算阶段又划出两个分阶段:前概念或象征思维阶段和直觉思维阶段。

(1) 前概念或象征思维阶段(2～4 岁)

这一阶段的产生标志是儿童开始运用象征符号。例如在游戏时,儿童用小木凳当汽车,用竹竿做马,木凳和竹竿是符号,而汽车和马则是符号象征的东西。即儿童已能够将这二者联系起来,凭着符号对客观事物加以象征化。客观事物(意义所指)的分化,皮亚杰认为就是思维的发生,同时意味着儿童的符号系统开始形成了。

语言实质上也是一种社会生活中产生并约定的象征符号。象征符号的创造及语言符号的掌握,使儿童的象征思维得到发展。但这时期的儿童语词只是语言符号附加上一些具体词,缺少一般性的概念,因而儿童常把某种个别现象生搬硬套到另一种现象之上,他们只能作特殊到特殊的传导推断,而不能作从一般到特殊的推理。从这个时期儿童常犯的一些错误可以看出这点。例如,儿童认识了牛,他也注意到牛是有四条腿的大动物,并且儿童已掌握"牛"。又如儿童看到别人有一顶与他同样的帽子,他会认为"这帽子是我的"。他们在房间看到一轮明月,而一会儿之后在马路上看到被云雾遮掩的月亮,便会认为天上有两个月亮。

(2) 直觉思维阶段(4～7 岁)

这一阶段是儿童智力由前概念思维向运算思维的过渡时期。此阶段儿童思维的显著特征是仍然缺乏守恒性和可逆性,但直觉思维开始由单维集中向二维集中过渡。守恒即将形成,运算思维就要到来。有人曾用两个不同年龄孩子挑选量多饮料的例子对此加以说明:一位父亲拿来两瓶可口可乐(这两瓶可口可乐瓶的大小形状一样,里面装的饮料也是等量),准备分别给他一个 6 岁和一个 8 岁的孩子,开始两孩子都知道两瓶中的饮料是一样多的。但父亲并没有直接将两瓶可乐饮料分配给孩子,而是将其中一瓶倒入了一个大杯中,另一瓶倒入了两个小杯中,再让两个孩子挑选。6 岁孩子先挑,他首先挑选了一大杯而放弃两小杯,可是当他拿起大杯看着两个小杯,又似乎犹豫起来,于是放下大杯又来到两小杯前,仍是拿不定主意,最后他还是拿了一大杯,并喃喃地说:"还是这杯多一点。"这个 6 岁的孩子在挑选饮料时表现出了犹豫地选择了大杯。在 6 岁孩子来回走动着挑选量较多的饮料时,他那 8 岁的哥哥却在一旁不耐烦而鄙薄地叫道:"笨蛋,两边是一样多的""如果你把可乐倒回瓶中,你就会知道两边是一样多的",他甚至还亲自示范了将饮料倒回瓶中以显示其正确性。从这个 6 岁孩子身上可以充分体现出直觉思维阶段儿童思维或智力的进步和局限性。数周

前毫不犹豫地挑选大杯说明他的思维是缺乏守恒性和可逆性的，他对量的多少的判断只注意到了杯子大这一个方面，而当他此次挑选过程中所表现出的迷惘则说明他不仅注意到了杯子的大小，也开始注意到杯子数量，直觉思维已开始从单维集中向两维集中过渡。但他最后挑选大杯表明守恒和可逆意识并未真正形成。

6 岁儿童挑选可乐过程表现出的迷惘和犹豫其实也是一种内心的冲突或不平衡，即同化与顺应之间的不平衡。过去的或是说现存的认知结构或图式(同化性认知结构)已不能解决当前问题，新的认知结构尚未建立。不平衡状态不能长期维持，这是智力的"适应"功能所决定的，平衡化因素将起作用，不平衡将向着平衡的方向发展，前运算阶段的认知结构将演变成具体运算思维的认知结构。守恒性和可逆性获得是这种结构演变的标志。8 岁男孩的叫喊和示范动作充分体现了这一点。

总结起来，前运算阶段的儿童认识活动有以下几个特点：①相对的具体性，借助于表象进行思维，还不能进行运算思维。②思维的不可逆性，缺乏守恒结构。③自我中心性，儿童站在自己经验的中心，只有参照他自己才能理解事物，他认识不到他的思维过程，缺乏一般性。他的谈话多半以自我为中心。④刻板性，表现为在思考眼前问题时，其注意力还不能转移，还不善于分配；在概括事物性质时缺乏等级的观念。

皮亚杰将此阶段的思维称为半逻辑思维，与感知运动阶段的无逻辑、无思维相比，这是一大进步。

### 3. 具体运算阶段(7～11 岁)

以儿童出现了内化了的、可逆的、有守恒前提的、有逻辑结构的动作为标志，儿童智力进入运算阶段，首先是具体运算阶段。

说运算是具体的运算意指儿童的思维运算必须有具体的事物支持，有些问题在具体事物帮助下可以顺利获得解决。皮亚杰举了这样的例子：爱迪丝的头发比苏珊淡一些，爱迪丝的头发比莉莎黑一些，问儿童："三个中谁的头发最黑"。这个问题若是以语言的形式出现，则具体运算阶段儿童难以正确回答。但如果拿来三个头发黑白程度不同的布娃娃，分别命名为爱迪丝、苏珊和莉莎，按题目的顺序两两拿出来给儿童看，儿童看过之后，提问者再将布娃娃收藏起来，再让儿童说谁的头发最黑，他们会毫无困难地指出苏珊的头发最黑。

具体运算阶段儿童智慧发展的最重要表现是获得了守恒性和可逆性的概念。守恒性包括有质量守恒、重量守恒、对应量守恒、面积守恒、体积守恒、长度守恒等。具体运算阶段儿童并不是同时获得这些守恒的，而是随着年龄的增长，先是在 7～8 岁获得质量守恒概念，之后是重量守恒(9～10 岁)、体积守恒(11～12 岁)。皮亚杰将质量守恒概念达到时作为儿童具体运算阶段的开始，而将体积守恒达到时作为具体运算阶段的终结或下一个运算阶段(形式运算阶段)的开始。

具体运算阶段儿童所获得的智慧成就有以下几个方面。

(1) 在可逆性(互反可逆性)形成的基础上，借助传递性，能够按照事物的某种性质如长短、大小、出现的时间先后进行顺序排列。例如给孩子一组棍子，长度(从长到短为 A、B、C、D……)相差不大。儿童会用系统的方法，先挑出其中最长的，然后依次挑出剩余棍子中最长的，逐步将棍子正确地顺序排列(这种顺序排列是一种运算能力)，即 A>B>C>D……当然孩子不会使用代数符号表示他的思维，但其能力实质是这样的。

(2) 产生了类的认识,获得了分类和包括的智慧动作。分类是按照某种性质来挑选事物,例如他们知道麻雀(用 A 表示)少于鸟(用 B 表示),鸟少于动物(C),动物少于生物(D),这既是一种分类包括能力,也是一种运算能力,即 A(麻雀)<B(鸟)<C(动物)<D(生物)。

(3) 把不同类的事物(互补的或非互补的)进行序列的对应。简单的对应形式为一一对应。例如给学生编号,一个学生对应于一个号,一个号也只能对应于一个学生,这便是一一对应。较复杂的对应有二重对应和多重对应。二重对应的例子,如一群人可以按肤色而且按国籍分类,每个人就有双重对应。

(4) 自我中心观进一步削弱,即去中心的,在感知运动阶段和前运算阶段,儿童是以自我为中心的,他以自己为参照系来看待每件事物,他的心理世界是唯一存在的心理世界,这妨碍了儿童客观地看待外部事物。在具体运算阶段,随着与外部世界的长期相互作用,自我中心逐渐克服。有研究者曾经做过这样一个实现:一个 6 岁的孩子(前运算阶段)和一个 8 岁的孩子(具体运算阶段)一起靠墙坐在一个有四面墙的房间里,墙的四面分别挂在区别明显的不同图案(A、B、C、D),同时这些图案被分别完整地拍摄下来制成四张照片(a、b、c、d)。让两个儿童先认真看看四面墙的图案,然后坐好,将四张照片显示在孩子面前,问两个儿童:哪一张照片显示的是你所靠坐墙对面的图案?两位孩子都困难地、正确地答出(a)。这时继续问孩子:假设你靠坐在那面墙坐,这四张照片中的那一张将显示你所靠坐墙(实际没有靠坐在那面墙、乃假设)对面的图案?6 岁的前运算阶段儿童仍然答的是他实际靠坐墙对面的图片(a),而 8 岁的具体运算阶段儿童指出了正确的图案照片(c)。为了使 6 岁的男孩对问题理解无误,研究者让 8 岁男孩坐到对面去,再问 6 岁孩子:8 岁孩子对面的墙的图案照片是哪一张?6 岁孩子仍然选了他自己靠坐墙对面的照片(a)。

概括起来,进入具体运算阶段的儿童获得了较系统的逻辑思维能力,包括思维的可逆性与守恒性;分类、顺序排列及对应能力,数的概念在运算水平上掌握(这使空间和时间的测量活动成为可能);自我中心观削弱等。

#### 4. 形式运算阶段(12～15 岁)

上面曾经谈到,具体运算阶段,儿童只能利用具体的事物、物体或过程来进行思维或运算,不能利用语言、文字陈述的事物和过程为基础来运算。例如爱迪丝、苏珊和莉莎头发谁黑的问题,具体运算阶段不能根据文字叙述来进行判断。而当儿童智力进入形式运算阶段,思维不必从具体事物和过程开始,可以利用语言文字,在头脑中想像和思维,重建事物和过程来解决问题。故儿童可以不很困难地答出苏珊的头发黑而不必借助于娃娃的具体形象。这种摆脱了具体事物束缚,利用语言文字在头脑中重建事物和过程来解决问题的运算就叫做形式运算。

除了利用语言文字外,形式运算阶段的儿童甚至可以根据概念、假设等为前提,进行假设演绎推理,得出结论。因此,形式运算也往往称为假设演绎运算。由于假设演泽思维是一切形式运算的基础,包括逻辑学、数学、自然科学和社会科学在内。因此儿童是否具有假设演绎运算能力是判断他智力高低的极其重要的尺度。

当然,处于形式运算阶段的儿童,不仅能进行假设演绎思维,皮亚杰认为他们还能够进行一切科学技术所需要的一些最基本运算。这些基本运算,除具体运算阶段的那些运算外,还包括这样的一些基本运算:考虑一切可能性;分离和控制变量,排除一切无关因素;观察

变量之间的函数关系,将有关原理组织成有机整体等。

形式运算思维是儿童智力发展的最高阶段。在此有两个问题应加以说明。

(1) 并非儿童成长到 12 岁以后就都具备形式运算思维水平,近些年在美国的研究发现,在美国大学生中(一般 18～22 岁),有约半数或更多的学生,其智力水平或仍处于具体运算阶段,或者处于具体运算和形式运算两个阶段之间的过渡。

(2) 15 岁以后人的智力还将继续发展,但总的来说属于形式运算水平。可以认为,形式运算阶段还可分出若干个阶段,有待进一步研究。皮亚杰认为智力的发展是受若干因素影响的,与年龄没有必然的联系。所以达到某一具体阶段的年龄即使有很大的差异并不构成皮亚杰理论的重大问题。

综上可知,在皮亚杰的发生认识论中运算思维结构是认识活动或智力活动的主要结构。他认为运算结构不仅是一种生物结构,而更重要的是一种逻辑结构。运算思维的基本特点是守恒性。所谓守恒,就是内化的、可逆的动作,守恒是通过逆反性和相互性实现的。

目前,国内外对儿童青少年思维的发展一般分为三个阶段:

(1) 直观(感知)行动思维;

(2) 具体形象思维;

(3) 抽象逻辑思维。

抽象逻辑思维又可以分为初步逻辑思维、经验型逻辑思维、理论型逻辑思维。

认识的问题是一个复杂的问题,每一个认识主体都处于复杂的社会联系之中,认识的产生和发展不可能不受到社会联系的制约。皮亚杰的发生认识论没有把儿童心理发展的研究放在社会联系之中并予以考察,因此也存在一定的缺陷。

### 11.4.3 新皮亚杰主义

新皮亚杰主义者是指这样一群研究者、理论家,他们持有与皮亚杰相同或相近的发展观,但在处理皮亚杰理论中的一些问题时,采取了某种更灵活的态度,采纳了许多其他理论的观念,特别是信息加工理论的一些观念。特定的材料、任务、社会信境及指导语似乎影响着儿童的表现,但皮亚杰从未对这种变化过程作出系统解释。

该学派的重要代表之一是凯斯(R. Case)的理论[128]。凯斯将认知变化视为对某个问题中越来越多的特征加以处理的过程。不同的发展水平,表现为达到最终目标而建立次级目标的不同能力水平;认知的发展,就是这类目标分化、协调和重新设定活动的不断重演。凯斯的认知发展理论把儿童比作问题解决者,认知发展好比是一系列功能不断强大的问题解决程序的更新,由此导致儿童理性认识的功效日益强大;这种发展外在地表现为达到次级目标和最终目标,儿童逐渐能够建构新的策略或利用合适的已有策略。凯斯还以儿童信息处理能力的增长来解释认知的发展,认为信息处理能力的增长是由神经系统的髓鞘化及与任务相关的运算实践所引起的。

新皮亚杰理论试图克服皮亚杰理论的不足,为认知发展寻求一种新的理论依据,以符合现代认知心理研究的总体趋势。他们把信息加工的观点与皮亚杰理论相结合,出现了三个新的趋势。

(1) 以研究儿童智力发展的共性转而解释其特殊形式,强调个体差异。他们认为儿童认知能力的发展并不是以阶段形式出现的,而是随个体知识和经验的增长不断发展。他们

提出的“风格”研究，引起了许多认知心理学家的兴趣。他们认为风格的本质就是适应、选择和塑造环境的不同方式。他们不但探讨个体如何形成“风格”，如何使“风格”导致某些态度和策略的偏爱，也探讨认知发展的普遍规律与个体变化之间的连接。这有助于帮助我们理解个体心理发展差异，重视社会环境(教育)的作用。

(2) 更加强调认知的情境性特征。情境观点认为，很大程度上知识依赖于背景，因此不能把它们独立于背景。因此在认知发展的不同领域，发展的模式和速率存在着个体间与个体内的差异。这一点与信息加工理论相符，为情境教学与情境实验提供了理论依据。

(3) 强调特定领域内的有关概念变化的过程与知识。随着时间的推移，儿童的知识逐渐独立于背景，但并不超出一定的学科领域。研究者更注重的是一般领域知识与特定领域知识之间的连接点。

近年来，新皮亚杰学派不仅在理论方面有了新的发展，而且在实践领域，特别是在教育实践领域也获得了日益广泛的应用。在婴儿教育方面，心理学工作者根据皮亚杰的感知运动阶段理论，指导婴儿摆弄物体，操作智力玩具等，帮助孩子形成对物体的特性(如色、形状、体积、质地等)的认识；在幼儿教育方面，设计了各种智力玩具和教具(如图片、积木等)，为儿童能提早形成数概念、空间概念及时间概念打下基础。

## 11.5　智力发展的影响因素

在儿童智力由低级向高级的演变过程中究竟有些什么影响因素呢？对这一问题传统上归为三个经典因素，即成熟、经验和社会环境。皮亚杰充分肯定这些因素在儿童智力发展的重要作用，认为这些因素是必不可少的。但他同时提出了第四种因素——不断成熟的内部组织和外部环境的相互作用因素，即平衡(又称调节)，并指出平衡化和自动调节是智力发展的决定因素。以下分别介绍这四个因素。

### 11.5.1　成熟因素

所谓成熟，即指在遗传程序控制下，机体、神经系统和内分泌系统逐渐发育成长的过程。在有的学者看来，儿童之所以随着年岁的增加而表现出心理和智慧的发展，乃是这种成熟的结果。即人的遗传基因型决定了其心理、智慧的发展水平，儿童的智力何时达到何种水平似乎早有安排，后天表现只不过是先天遗传因素的逐渐显露。这种观点的极端形式是“遗传决定论”，代表人物是优生学创始人英国的高尔顿(F. Galton)。而“自然成熟论”(代表人物是盖塞尔)虽也不忽视环境因素的作用，但始终认为儿童智能的发展有一定的生物内在进度表。

皮亚杰认为神经系统的成熟对智力发展有着重要作用。因为智力作为人类的一种高级机能，它必然依赖于一定的神经及内分泌系统的生理基础。因此生理机能的成熟无疑就成为智慧发展的必要因素。这种成熟因素在使儿童心理及智慧的发展遵循不变的连续的阶段方面起着不可缺少的作用。例如新生婴儿的吸吮反射、拥抱反射的生理基础是反射弧，无反射便不可能有这些反射。当神经系统的锥体束中的神经纤维髓鞘化后(相当于婴儿 4 个半月)，婴儿便有了视觉与抓握反射的协调(感知运动阶段的第三分阶段)。但皮亚杰认为智力

的成长过程中，成熟不是决定条件，神经系统的成熟只能决定某一给定阶段的可能性与不可能性。环境因素对于实现这些可能性是始终不可少的。可以这样认为，即使在心理或智力发展的初级阶段，一些简单的初级心理机能（如感知、动作以及初始言语），虽然遗传成熟的制约因素较大，但亦需要最低限度的习得经验和机能练习。而一些较复杂的高级心理机能的获得与发展，则更多地是受环境因素与机体成熟因素之间动态交互影响的结果。皮亚杰说："我们不能设想有一种作为人类智力发展基础的遗传程序存在。"成熟不能说明计算 2＋2＝4 的能力和演绎推理是如何形成的。智力不是天生的。概念也不是天生的，就拿与年龄有密切关系的语言来说，如果一个儿童不处于人类社会中，就不会在任何年龄获得人类语言。

概括起来，成熟是影响智力发展的一个因素，它为智力结构的演化提供了可能，但是在可能性和现实性之间，还必须有一些其他因素，例如练习、经验和社会的相互作用。

## 11.5.2　经验因素

传统上说明认知发展的第二个因素是经验。皮亚杰认为经验对人的智力发展是不可缺少的。经验因素包括物理环境和自然环境。他非常重视经验，指出经验是知识的来源，是智力增长的重要条件，但是经验因素也是不充分的，不能决定心理及智慧的发展。所谓物理经验是通过一种简单的抽象过程从客体本身中引出的。例如儿童关于物体的重量、物体的颜色、物体表面的光滑程度、声音的高低、木块浮在水面、水结成冰等经验是通过儿童的触觉、视觉、听觉等从上述物体中抽出来的。这种经验最本质的特点是来源于物体本身，这些物体的性质（重量大小、声音高低）是客观存在的，即使儿童不去看、不去摸或不去作用于这些物体，这些物体的性质依然存在。

逻辑数理经验虽也来源于主体与客体的相互作用中，但这种经验不是由物体抽出，而是产生于主体客体所施的动作及协调。皮亚杰举过一个例子解释这种逻辑数理经验：他有一位数学家朋友，小时在沙滩上玩卵石，他把 10 个卵石排成一行，发现不论从那端开始数都是 10 个，然后他又把卵石排成另外的形状，如排成圆形、四方形，数出来的数目仍然不变。于是他得出"和与顺序无关"的结论。皮亚杰认为，这件事对于成人来说极为平常，但对儿童来说却是一件了不起的发现。在玩卵石的时候，可以感受到卵石的重量、形状及大小等，这是物理经验。而"和与顺序无关"也是经验，但它不是由感知的直觉获得，反映的也不是卵石的物理性状，故不是物理经验。儿童是通过计数卵石的动作得到的这种经验，它是关于数和数的交换性的概念，这就是逻辑数理经验。

物理经验和逻辑数理经验是本质上完全不同的两种经验。由物理经验可以认识物理性质，但物理性质不依赖于物理经验。有物理性质无物理经验（无动作）也存在；逻辑数理经验来源于动作，而不依赖于物理性质，无动作则无逻辑经验来源于动作，而不依赖物体的物理性质，无动作则无逻辑数理经验。两种经验包含着性质不同的两种抽象过程，物理经验是一种简单的本义的抽象，只考虑物体某一性质（如重量），不考虑其他，即只把"重量"抽象出来；而逻辑数理经验是一种反省的抽象，这种抽象由于是对自身动作的抽象，这就不仅要求不考虑其他特性，还需要一个新的再建过程。

前面介绍过，一切运算都是动作，所形成的经验都是逻辑数理经验。逻辑数理经验对于认知结构的形成有极其重要的意义，智力主要表现在具有最必要的逻辑数理经验。任何一

个动作都可以抽出物理经验和逻辑数理经验,但人们一般容易注意获得物理经验而不容易注意获得逻辑数理经验。因而在儿童智力培养中,一方面应该注意丰富儿童的生活,提供各种的自然环境材料,使儿童获得物理经验;在另一方面,也许是更重要的,应该在上述活动环境中,指导孩子通过分析、综合、思索和探究事物之间的内在联系和规律,获得逻辑数理经验。

但皮亚杰提出两点理由认为经验不能说明一切,也不是儿童智力发展的决定因素。

(1) 有些概念不能从经验中抽出,即概念不完全取决于经验。例如儿童往往先获得质量守恒概念,而后获得重量及体积守恒概念,重量和体积通过儿童对物体测量得以理解,但儿童没有通过经验获得守恒概念。皮亚杰问,在尚无重量守恒和体积守恒时,质量守恒概念从何处来?

(2) 经验这一概念是有歧义和含糊的。

### 11.5.3 社会环境因素

社会环境因素主要涉及社会生活、教育、学习及语言等方面。很显然,这些因素对儿童智能发展的作用是巨大的。首先是社会生活。人的一生就浸润于社会生活环境中,婴儿自出生的一刹那起就开始了其社会化的一生。社会生活对儿童智力发展的影响是明显的。例如有人研究就发现,儿童在2岁前与父母亲之间的关系(或称家庭情感气氛)与孩子长到18岁时的智力是呈正相关的,关系较好,智力较高;否则就低。另外从儿童"自我中心观"的发展也可以看出,随着儿童与家庭成员及小伙伴、老师之间的接触,儿童将出现"去中心化",这也是儿童智力发展的一个表现。其次是教育,事实上当强调经验因素在促进儿童思维发展中的作用时,实际上已蕴育着对教育因素的重视。因为系统的教育(学习和训练)可以使儿童更好地感受外部世界,获得经验(包括物理经验和逻辑数理经验),也正是因为这一点,教育因素才能促进儿童智慧的发展。

皮亚杰十分强调教育必须符合于儿童的认知结构。他说:"即使在主体似乎非常被动的社会传递例如学校教育的情况下,如果缺少儿童的主动同化作用,这种社会作用无效,而儿童主动的同化作用则是以适当的运算结构为前提的"。又说:"只是当所教的东西可以引起儿童积极从事再造和再创的活动,才能有效地被儿童所同化"。教育可在一定程度上加速儿童智力发展阶段的过渡,但并不能超越或改变发展的顺序,任何儿童(包括天才),也是绝无例外的。在对儿童实施早期教育时,应充分重视这一点。

再谈语言,语言在动作内化于表象和思想方面起主导作用,在介绍前运算阶段儿童的思维发展中,已经可看出语言对儿童智力的作用。但语言不是唯一起作用的因素,语言是一种符号系统。但它不是唯一的符号系统。尽管它可能是最佳的符号系统,属于符号系统之列的还有:图画、造型、模仿动作、内化的模仿、特异的手势和姿态等。虽然语言与智力关系密切,但两者的发展并不平行,有的人语言流畅,但智力平平;而有的人思维能力极其优异,却可能不善于辞令。

可以说没有社会传递就不会有人类社会的全部科学文化的继承和发展。皮亚杰十分强调社会环境因素在儿童智力发展中的重要作用。但是儿童智力发展具有连续性这一事实又说明社会环境因素不是发展的充分因素。从上述对语言、教育等的分析中也看出这一点。

### 11.5.4　平衡化因素

生理成熟、自然环境和社会环境都是儿童智慧发展必不可少的前提和条件。然而各自都不是充分的条件。儿童智慧成长也不是这些因素简单机械相加的结果。皮亚杰提出了平衡化因素的概念并认为平衡化是儿童智力发展的决定因素。在皮亚杰看来,既然成熟、经验及社会环境各自都不能完全解释发展的根本原因,那么必然存在其他因素,这个因素在原有三种因素之间起着协调或调节作用,这个协调者或调节者就是平衡化。

皮亚杰的智力观前已述及,他认为“智慧是生物适应性的一种特殊表现”。智力是一切认知结构趋于平衡的形式之一。智力是有结构基础的,智力的提高就是智力结构的不断发展,同化与顺应是智力不断建构发展的两个基本过程。个体在遇到外部刺激(自然环境与社会环境)时,首先与之发生作用的是现存的图式,这种图式,在婴儿初生时,是遗传决定的先天图式,这个先天图式,随着成熟及成熟的机体与外界的交互作用而逐渐演变成现存图式。按照现存图式,机体吸收外界的信息并做出反应,此即是同化过程,反复的同化使图式或认知结构得到巩固。人们在认识事物,解决问题时总是利用原有的思维和行为模式,这就是同化的表现。当一个新刺激到来,机体仍用原有的或现存的图式去应付,但结果可能是不成功的(如前运算阶段向具体运算过滤时期儿童挑选可乐饮料时感到迷惘和犹豫),于是新刺激在被主体同化的认知格式吸纳的同时,将使这一认知格式发生改变,即为顺应。改变同化性认知格式并不是瞬间完成的,而需经历一定的过程。过程进行中同化与顺应处于一种不平衡状态,旧的图式与将形成的新图式之间存在冲突,表现在人们认知方面即是旧观念与新观念的斗争。如果新图式终于建成,即宣告同化与顺应的不平衡状态结束,平衡已经实现,儿童智力获得了发展。以此新图式为基础,儿童又开始了新的同化,同化中建构,又遇到新刺激,出现新的顺应,儿童智力正是这样一步一步由低级向高级发展的。成熟、自然环境和社会环境都在发展中起作用,而平衡化因素则调节着这三个因素使儿童智力向着一定的方向发展。

## 11.6　智力发展的人工系统

随着计算机科学技术的发展,人们试图通过计算机或其他人工系统对于生物学的机理进行深入的理解,用计算机复制自然和自然生命的现象和行为,于 1987 年建立了人工生命的新学科[403]。人工生命是指用计算机和精密机械等生成或构造表现自然生命系统行为特点的仿真系统或模型系统,体现自然生命系统的组织和行为过程,自然生命系统的行为特点和动力学原则表现为自组织、自修复、自复制的基本性质,以及形成这些性质的混沌动力学,环境适应性及其进化[936]。

研究人工生命的智力发展,使人工生命也像人一样通过自主学习变得越来越聪明。最根本的或者说是最本质的问题是:开发人工生命像人一样的学习能力。这是机器智能研究的一个巨大挑战。在过去的几十年里,人们主要采用四种方法来研究机器智力发育[802]。

(1) 基于知识的方法。对机器进行直接编程从而完成预定的任务。

(2) 基于行为的方法。用行为模型来取代传统的世界模型,智能程序开发者针对不同

层次的行为状态和所期望的行为编写程序。这种方法的特点是基于行为的手动建模和基于行为的手动编码。

(3) 遗传搜索方法。在计算机模拟的虚拟世界中,机器按照适者生存的原则进化。但没有一种方法使得机器能像成年人一样,具有处理复杂、多变事务的综合能力。

(4) 基于学习的方法。机器在具体任务学习程序的控制下,输入人类编辑好的感知数据,如有教师学习和强化学习。但由于学习过程是非自动的,训练系统时的开销比较大。

传统手工机器智能开发的具体过程是:首先让人类专家弄清楚所需求解问题(或任务)的具体内容,接着由人类专家根据具体问题设计其知识表示方法,然后利用设计好的知识表示进行具体问题的程序设计,最后运行所谓的"智能"程序。在程序执行的过程中,如果利用感知数据对上述预先设计的知识及有关参数进行修改,这就是机器学习。在传统机器智能开发方式下,机器只会做事先设计好的事情。事实上,机器根本搞不清自己在做什么。

自主机器智力开发程式不同于传统的机器智能开发程式[802],主要包含下列内容:首先根据机器的生态工作条件(如陆地、水下等环境)设计合适的机器,然后在此基础上设计机器智力开发程序,并在机器投入使用时(或者说"出生"时)运行机器智力开发程序。为达到开发机器智力之目标,人类需要不断地与机器实时交互来培养正在进行智力开发的机器。由此可见,机器的智力发育也是一个漫长的过程,其本质是使机器自主地生活并使它越来越聪明。

我们将自主学习机制引入智能体(Agent),目标是为了让智能体具有像人类一样的自主学习能力,其结构如图11.2所示[937]。其中控制中枢和自主智力发展(Autonomous Mental Development,AMD)是智能体的根本,知识库、通信机制、感知器和效应器也是一个具有自主学习能力的智能体的必不可少的组件。控制中枢类似于人脑的神经中枢,对其他各组件起控制和协调作用,反应智能体的功能也在控制中枢中得到体现。AMD是智能体的自主学习系统,其功能体现为一个智能体的自主学习能力。通信机制采用通信语言(如ACL)直接与智能体所处的环境进行信息交互,它是一个特殊的感知器或效应器。感知器就如同人的眼睛和耳朵等感觉器官,用于感知智能体所处的环境。效应器就如同人的手脚、嘴等器官,用于完成智能体所要做的事情。智能体通过执行AMD的AA学习(英文全称为Automated Animal—like learning)算法不断地增长自己的知识,提高自己的能力,主要体现在功能模块数量的不断增加和功能的不断增强上。知识库相当于人的大脑的记忆部件,用于存储信息。自主机器智力开发的一个非常重要的功能就是信息的自动存储,因此如何有效地自动组织并存储各种类型的信息(如图像、声音、文本等)是一个AMD成功的关键。

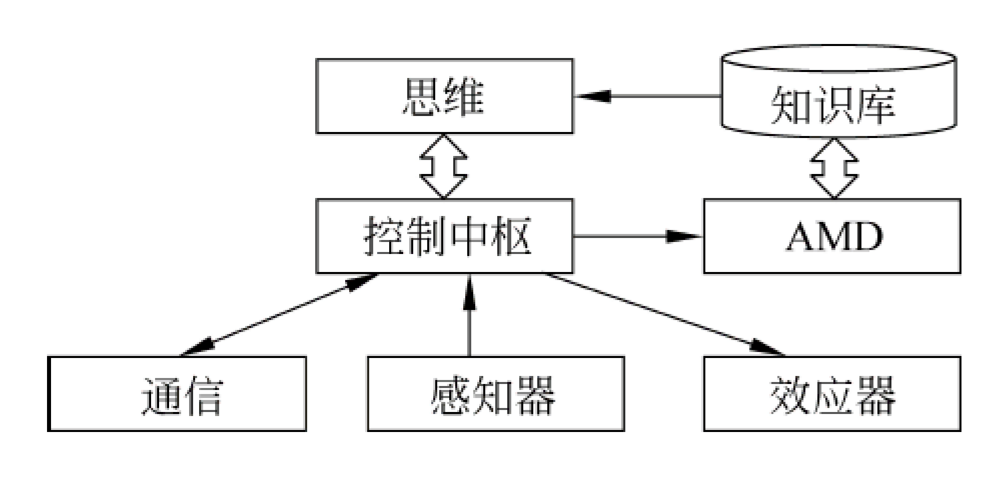



图11.2 自主智力发展

# 第 12 章<<
# 情绪与情感

情绪是对外界事物态度的主观体验,是人脑对客观外界事物与主体需求之间关系的反应,是多种感觉、思想和行为综合产生的心理和生理状态。在智能科学研究中,要想真正的或者更大程度上模拟真实的人类高级功能,还必须深入考虑情感因素的作用。机器智能只有被赋予了情感的成分,才能实现有效的人机交互。

## 12.1 概述

人类在认识外界事物时,会产生喜与悲、乐与苦、爱与恨等主观体验。我们把人对客观事物的态度体验及相应的行为反应称为情绪。这里,概要介绍情绪的构成要素、基本形式和功能。

### 12.1.1 情绪的构成要素

情绪的构成包括三种层面：在认知层面上的主观体验,在生理层面上的生理唤醒,在表达层面上的外部行为。当情绪产生时,这三种层面共同活动,构成一个完整的情绪体验过程。

#### 1. 主观体验

情绪的主观体验是人的一种自我觉察,即大脑的一种感受状态。人有许多主观感受,如喜、怒、哀、乐、爱、惧、恨等。人们对事物的态度不同会产生不同的感受。人对自己、对他人、对事物都会产生一定的态度,如对朋友遭遇的同情,对敌人凶暴的仇恨,事业成功的欢乐,考试失败的悲伤。这些主观体验只有个人内心才能真正感受到或意识到,如我知道“我很高兴”,我意识到“我很痛苦”,我感受到“我很内疚”,等等。

#### 2. 生理唤醒

生理唤醒是指情绪与情感产生的生理反应。它涉及广泛的神经结构,如中枢神经系统的脑干、中央灰质、丘脑、杏仁核、下丘脑、蓝斑、松果体、前额皮层,及外周神经系统和内、外分泌腺等。生理唤醒是一种生理的激活水平。不同情绪、情感的生理反应模式是不一样的,如满意、愉快时心跳节律正常；恐惧或暴怒时,心跳加速、血压升高、呼吸频率增加甚至出现间歇或停顿；痛苦时血管容积缩小等。脉搏加快、肌肉紧张、血压升高及血流加快等生理指数,是一种内部的生理反应过程,常常是伴随不同情绪产生的。

#### 3. 外部行为

在情绪产生时,人们还会出现一些外部反应过程,这一过程也是情绪的表达过程。如人

悲伤时会痛哭流涕,激动时会手舞足蹈,高兴时会开怀大笑。情绪所伴随出现的这些相应的身体姿态和面部表情,就是情绪的外部行为。它经常成为人们判断和推测情绪的外部指标。但由于人类心理的复杂性,有时人们的外部行为会出现与主观体验不一致的现象。比如在一大群人面前演讲时,明明心里非常紧张,还要做出镇定自若的样子。

主观体验、生理唤醒和外部行为作为情绪的三个组成部分,在评定情绪时缺一不可,只有三者同时活动,同时存在,才能构成一个完整的情绪体验过程。例如,当一个人佯装愤怒时,他只是愤怒的外在行为,却没有真正的内在主观体验和生理唤醒,因而也就称不上有真正的情绪过程。因此,情绪必须是上述三方面同时存在,并且有一一对应的关系,一旦出现不对应,便无法确定真正的情绪是什么。这也正是情绪研究的复杂性,以及对情绪下定义的困难所在。

在现实生活中,情绪与情感是紧密联系在一起的,但二者却存在着一些差异。

(1) 从需要的角度看差异。

情绪更多地是与人的物质或生理需要相联系的态度体验。如当人们满足了饥渴需要时会感到高兴,当人们的生命安全受到威胁时会感到恐惧,这些都是人的情绪反应。情感更多地与人的精神或社会需要相联系。如友谊感的产生是由于我们的交往需要得到了满足,当人们获得成功时会产生成就感。友谊感和成就感就是情感。

(2) 从发生早晚的角度看差异。

从发展的角度来看,情绪发生早,情感产生晚。人出生时会有情绪反应,但没有情感。情绪是人与动物所共有的,而情感是人所特有的,它是随着人的年龄增长而逐渐发展起来的。如人刚生下来时,并没有道德感、成就感和美感等,这些情感反应是随着儿童的社会化过程而逐渐形成的。

(3) 从反映特点看差异。

情绪与情感的反映特点不同。情绪具有情境性、激动性、暂时性、表浅性与外显性,如当我们遇到危险时会极度恐惧,但危险过后恐惧会消失。情感具有稳定性、持久性、深刻性、内隐性,如大多数人不论遇到什么挫折,其民族自尊心不会轻易改变。父辈对下一代殷切的期望、深沉的爱都体现了情感的深刻性与内隐性。

实际上,情绪和情感既有区别又有联系,它们总是彼此依存,相互交融在一起。稳定的情感是在情绪的基础上形成起来的,同时又通过情绪反应得以表达,因此离开情绪的情感是不存在的。而情绪的变化也往往反映了情感的深度,而且在情绪变化的过程中,常常包含着情感。

### 12.1.2 情绪的基本形式

人类具有四种基本的情绪:快乐、愤怒、恐惧和悲哀。快乐是一种追求并达到目的时所产生的满足体验。它是具有正性享乐色调的情绪,具有较高的享乐维和确信维,使人产生超越感、自由感和接纳感。愤怒是由于受到干扰而使人不能达到目标时所产生的体验。当人们意识到某些不合理的或充满恶意的因素存在时,愤怒会骤然发生。恐惧是企图摆脱、逃避某种危险情景时所产生的体验。引起恐惧的重要原因是缺乏处理可怕情景的能力与手段。悲哀是在失去心爱的对象或愿望破灭、理想不能实现时所产生的体验。悲哀情绪体验的程度取决于对象、愿望、理想的重要性与价值。

在以上四种基本情绪之上，可以派生出众多的复杂情绪，如厌恶、羞耻、悔恨、嫉妒、喜欢、同情等。

## 12.1.3　情绪状态

依据情绪发生的强度、速度、紧张度、持续性等指标，可将情绪分为心境、激情和应激。

### 1. 心境

心境是一种具有感染性的、比较平稳而持久的情绪状态。当人处于某种心境时，会以同样的情绪体验看待周围事物。如人伤感时，会见花落泪，对月伤怀。心境体现了“忧者见之则忧，喜者见之则喜”的弥散性特点。平稳的心境可持续几小时、几周或几个月，甚至一年以上。

### 2. 激情

激情是一种爆发快、强烈而短暂的情绪体验。如在突如其来的外在刺激作用下，人会产生勃然大怒、暴跳如雷、欣喜若狂等情绪反应。在这样的激情状态下，人的外部行为表现比较明显，生理的唤醒程度也较高，因而很容易失去理智，甚至做出不顾一切的鲁莽行为。因此，在激情状态下，要注意调控自己的情绪，以避免冲动性行为。

### 3. 应激

应激是指在意外的紧急情况下所产生的适应性反应。当人面临危险或突发事件时，人的身心会处于高度紧张状态，引发一系列生理反应，如肌肉紧张、心率加快、呼吸变快、血压升高、血糖增高等。例如，当遭遇歹徒抢劫时，人就可能会产生上述的生理反应，从而积聚力量以进行反抗。但应激的状态不能维持过久，因为这样很消耗人的体力和心理能量。若长时间处于应激状态，可能导致适应性疾病的发生。

## 12.1.4　情绪的功能

### 1. 情绪的动机作用

情绪与动机的关系十分密切，情绪能够以一种与生理性动机或社会性动机相同的方式激发和引导行为。有时我们会努力去做某件事，只因为这件事能够给我们带来愉快与喜悦。从情绪的动力性特征看，分为积极增力的情绪和消极减力的情绪。快乐、热爱、自信等积极增力的情绪会提高人们的活动能力，而恐惧、痛苦、自卑等消极减力的情绪则会降低人们活动的积极性。有些情绪同时兼具增力与减力两种动力性质，如悲痛可以使人消沉，也可以使人化悲痛为力量。

情绪也可能与动机引发的行为同时出现，情绪的表达能够直接反映个体内在动机的强度与方向。所以，情绪也被视为动机潜力分析的指标，即对动机的认识可以通过对情绪的辨别与分析来实现。动机潜力是在具有挑战性环境下所表现出的行为变化能力。例如，当个体面对一个危险的情境时，动机潜力会发生作用，促使个体做出应激的行为。对这个动机潜力的分析可以由对情绪的分析获得。当面对应激场面时，个体的情绪会发生生理的、体验的

以及行为的三方面的变化,这些变化会告诉我们个体在应激场合动机潜力的方向和强度。当面临危险时,有的人头脑清晰,沉着冷静地离开;而有些人则惊慌失措,浑身发抖,不能有效地逃离现场。这些情绪指标可以反映出人们动机潜能的个体差异。

### 2. 情绪是心理活动的组织者

情绪对认知活动的作用,只用“驱动”来描述是不够的,情绪可以调节认知的加工过程和人的行为。诸如情绪自身的操作可以影响知觉中对信息的选择,监视信息的流动,因此情绪可以驾驭行为,支配有机体同环境相协调,使有机体对环境信息作出最佳处理。同时,认知加工对信息的评价通过神经激活而诱导情绪。在这样的相互作用中,无论情绪或认知,作为心理的东西,都以其内容而起作用。所不同的是,认知是以外界情境事件本身的意义而起作用;而情绪则以情境事件对有机体的意义,通过体验快乐或悲伤、愤怒或恐惧而起作用。它们之间的根本性质上的区别所导致的后果,在于情绪具备动机的作用而能激活有机体的能量,从而制约认知和行动。就此而言,情绪似乎是脑内的一个监测系统,调节着其他的心理过程。

近年来,情绪心理学家把情绪对其他心理过程的作用具体化为组织作用。其含义包括着组织的功能和破坏的功能。一般来说,正情绪起协调的、组织的作用,而负情绪起破坏的、瓦解的或阻断的作用。叶克斯-道森规律标示情绪在不同唤醒水平对手工操作的效果有所不同,而呈现为一个倒“U”字模式。

### 3. 情绪的健康功能

人对社会的适应是通过调节情绪来进行的,情绪调控的好坏会直接影响到身心健康。常听人们叹息“人生苦短”,在一般人的情绪生活中,常是苦多于乐。在喜怒哀乐爱惧恨中,正面情绪占 3/7,反面情绪占 4/7。情绪对健康的影响作用是众所周知的。积极的情绪有助于身心健康,消极的情绪会引起人的各种疾病。我国古代医书《内经》中就有“怒伤肝,喜伤心,思伤脾,忧伤肺,恐伤肾”的记载。有许多疾病与人的情绪失调有关,如溃疡、偏头痛、高血压、哮喘、月经失调等。有些人患癌症也与长期心情压抑有关。一项长达 30 年的关于情绪与健康关系的追踪研究发现,年轻时性情压抑、焦虑和愤怒的人患结核病、心脏病和癌症的比例是性情沉稳的人的 4 倍。所以,积极而正常的情绪体验是保持心理平衡与身体健康的条件。曾有人说过,一个小丑进城胜过一打医生,就非常形象地说明了情绪对人身体健康的影响。

### 4. 情绪的信号功能

情绪是人们社会交往中的一种心理表现形式。情绪的外部表现是表情,表情具有信号传递作用,属于一种非言语性交际。人们可以凭借一定的表情来传递情感信息和思想愿望。心理学家研究了英语使用者的交往现象后发现,在日常生活中,55%的信息是靠非言语表情传递的,38%的信息是靠言语表情传递的,只有 7%的信息才是靠言语传递的。表情是比言语产生更早的心理现象,在婴儿不会说话之前,主要是靠表情来与他人交流的。表情比语言更具生动性、表现力、神秘性和敏感性。特别是在言语信息暧昧不清时,表情往往具有补充作用,人们可以通过表情准确而微妙地表达自己的思想感情,也可以通过表情去辨认对方的

态度和内心世界。所以,表情作为情感交流的一种方式,它被视为人际关系的纽带。

## 12.2　情绪理论

### 12.2.1　詹姆斯-兰格情绪学说

19 世纪的美国心理学家威廉·詹姆斯(W. James)和丹麦生理学家卡尔·兰格(C. Lange)分别于 1884 年和 1885 年提出了相似的情绪理论。该理论基于情绪状态和生理变化的直接联系,提出情绪是对机体变化的感知,是机体各种器官变化时所引起的感觉的总和。詹姆斯说:"我认为,当我们一知觉到使我们激动的对象时,立刻就引起身体上的变化。在这些变化出现之时,我们对这些变化的感觉,就是情绪。""我们因为哭,所以愁;因为动手打人,才生气;因为发抖,所以怕。并不是我们愁了才哭,生气了才打,怕了才发抖。"兰格认为:"任何作用凡能引起广泛的血管神经系统功能的变化的,都会有情绪表现。"詹姆斯·兰格情绪学说强调生理变化对情绪的作用,有一定的历史意义,但它片面夸大了外围性变化对情绪的作用,而忽略了中枢对情绪的主导作用。

### 12.2.2　情绪评估-兴奋学说

美国心理学家阿诺德(M. B. Arnold)在 20 世纪 50 年代提出的情绪评估-兴奋学说,强调来自外界环境的影响要经过人的评价与估量才产生情绪,这种评价与估量是在大脑皮层上产生的。例如在森林里看到一只熊会引起惧怕,但在动物园里看到一只关在笼子里的熊却并不惧怕,这就是个体对情境的认识和评价在起作用。阿诺德给情绪下的定义是:情绪是驱利避害的一种体验倾向。他认为情绪反应包括机体内部器官和骨骼肌的变化,也认为对外围变化的反馈是情绪的基础。阿诺德认为皮质兴奋是情绪的主要机制。

### 12.2.3　情绪三因素说

20 世纪 70 年代,美国心理学家沙赫特(S. Schachter)提出了情绪三因素说,认为情绪的产生不是单纯地决定于外界刺激和机体内部的生理变化,而把情绪的产生归因于三个因素的综合作用,即刺激因素、生理因素和认知因素。他认为,认知因素中对当前情境的估计和过去经验的回忆在情绪形成中起着重要作用。例如,某人在过去经验中遇到的某种危险的情境,但能平安度过,当他再次经历这种险境时,回忆起过去的经验,便能泰然自若。也就是说,当现实情境与过去建立的经验模式相一致,相信能加以应付,人就没有明显情绪;当现实情境与预期和愿望不一致,感到无力应付时,就会产生紧张情绪。这种学说更加强调人的认知过程对情绪的调控作用。

### 12.2.4　基本情绪论

基本情绪论认为情绪在发生上有原型形式,即存在着数种泛人类的基本情绪类型,每种类型各有其独特的体验特性、生理唤醒模式和外显模式,其不同形式的组合形成了所有的人类情绪。从个体发展角度来看,基本情绪的产生是有机体自然成熟的结果,而不是习得的。

从生物进化的观点看,情绪原型是适应和进化的产物,也是适应和进化的手段;从猿到人类的进化、古皮质到新皮质的发展,面部肌肉系统的分化和面部血管的分布,以及情绪的发生和分化,都是同步进行和获得的。对于有哪些基本情绪则有不同看法,最常被提到的是厌恶、愤怒、高兴、悲伤、害怕等。

支持基本情绪理论的最著名的研究是埃克曼(Ekman)和伊扎德(Izard)进行的面部表情和运动反应的研究。埃克曼等要求新的被试者设想自己是某个故事情节中的人物,并且尽可能的表现出故事中人物的面部表情;同时研究人员对他们的面部表情进行了录像;最后让美国学生观看这些表情,并要求进行识别,结果美国学生能够从六种表情中识别出四种(快乐、愤怒、厌恶、悲伤)。孟昭兰等的实验也证明,中国婴儿和西方标准化基本情绪表情模式是一致的;同样,中国婴儿同中国成年人的基本情绪的表情模式也是一致的,社会化了的成人表情中仍然保留着基本表情模式。列文森(Levenson)等以西苏门答腊的年轻人作为被试,指导他们运动面部特定肌肉以外显基本情绪,并进行一系列生理学测量,最后把测量结果与美国大学生被试比较,结果发现与基本情绪相联系的自主神经系统的生理反应模式具有很大的跨文化一致性,这样的实验结果趋向于证实各基本情绪存在着泛人类的特定生理唤醒模式。

对基本情绪论的反对意见主要有:

(1) 尽管上述实验证实了情绪的泛文化模式,但这方面的研究也显示了某些跨文化的差异。

(2) 心理学和语言学研究已经发现,基本情绪词的含义在不同语言之间具有显著差异。

(3) 没有足够证据证明各基本情绪具有各自不同的神经生理机制。

(4) 面部表情识别的基础也许并不是基本情绪类型,而是面部表情在情绪体验的双极维度上的位置,或者是面部表情诱发的行为预备模式等。

### 12.2.5 维度论

维度论认为几个维度组成的空间包括了人类所有的情绪。维度论把不同情绪看作是逐渐的、平稳的转变,不同情绪之间的相似性和差异性是根据彼此在维度空间中的距离来显示的。最近二十年,维度途径得到了许多研究人员的青睐,但是对采用哪些维度也有许多争论。最广为接受的维度模式是如下两个维度组成的二维空间:①效价(Valence)或者愉悦度(Hedonic Tone),其理论基础是正负情绪的分离激活;②唤醒度(Arousal)或者激活度(Activation),指与情感状态相联系的机体能量激活的程度,唤醒的作用是调动机体的机能,为行动做准备。目前大家已经倾向于把激活维度和综合生理唤醒或者情绪体验的强度相联系。早在20世纪70年代,曼德勒(Mandle)受到信息加工理论的影响,提出自主性唤醒的知觉决定情绪体验的强度,认知评价决定情绪的性质,两者的整合上升到意识,从而产生情绪体验。国际情绪图片系统(IAPS)很好地体现了上述二维空间,在愉悦度和唤醒度构成的二维坐标空间中,被试者对情绪图片的平均评定值呈现规律性分布。

脑成像研究证实了正性和负性情绪的分离,即两者各自具有特定的大脑加工系统,分别与左半球和右半球活动相联系,或者说分别是左半球和右半球优势。电生理学研究表明,电影诱发厌恶和恐惧时,右侧额叶和颞叶活动增强,而正性情绪时则显示左半球相应脑区的活动增强。PET 研究也显示了类似的结果,给被试者呈现情绪图片诱发其特定情绪,结果发

现负性情绪时右额叶等脑区的代谢率增加，正性情绪时则左侧占优势。

勃莱德列(Bradley)等给被试者观看图片(选自 IAPS)，同时测量生理反应，以及在每次图片呈现后用自我评价模式去获得愉悦度、激活度、优势度的等级评定。实验结果发现，尽管并不是所有的生理反应与情感自我评价发生一致的变化，但总的来说，情绪的生理反应模式与情绪自我报告的二维模式是一致的。比如，惊反射(Startle Reflex)与情绪评价的效价密切相关：在正性效价，惊反射强度随着快乐程度的增加而下降，对于负性效价，惊反射强度随着负性程度的增加而增加；而皮肤的电反应与唤醒度一致：随着唤醒度的下降，皮肤的导电性也下降。这在某种程度上说明了惊反射是愉悦度的生理指标，皮电反应是唤醒度的生理指标。

愉悦度/唤醒度二维情绪空间的局限是并不能区分所有的情绪，比如对于同样是高唤醒低效价的愤怒和恐惧就无法明确区分，两者在空间中的位置相当接近。最近几年，人们开始热衷于用趋近/逃避(Approach-Withdrawal)来代替愉悦度。趋避度的优越性在于：①它更具有生物学基础，所有生物对于环境事件的行为反应都可以用趋避性来表示，而行为反应又是与情绪反应紧密相关的；②趋避度能够对于愉悦度无法区分的愤怒和恐惧作出区分，愤怒与正情绪一样导致趋近事物，恐惧导致逃避。愤怒和正性情绪都与趋近行为有关，这使人感到困惑，因为愤怒是一种负性情绪。然而我们必须记住，当面临一个有害事件，与愤怒有关的趋近会导致进攻行为以祛除有害事物，这符合生物适应意义。与愤怒属于趋近维度一致，哈门(Harmon)等的研究证明愤怒与左侧前额叶的活动增强有关。但是趋避维度的反对者认为，趋近维度和快乐维度的差异也说明行为趋近系统并不完全与正性情绪相关，情绪和行为是两回事。从狭义的情绪定义来说，趋避行为不属于情绪范畴，起码它一定不属于情绪体验的范畴，而是情绪体验后发生的动机和行为趋向。正性情绪不一定导致趋近行为，例如，如果爱夹杂着害羞或者不安全感，就导致逃避行为；负性情绪也不一定导致逃避行为，例如，尽管看恐怖电影会令人产生恐惧和厌恶，但这并不妨碍许多人对此的热衷；而对于中性或者不熟悉的刺激，更可能会采取轻微的趋近行为，这有利于探索未知世界，也符合自然生存法则。对于维度途径的不同意见主要有：

(1) 情绪评价具有个体差异，某些人以维度的方式去感受并报告情绪，而有些人体验和报告情绪的方式则更符合基本情绪理论。

(2) 尽管一些研究证实正负情绪的半球差异，但是也有一些研究却并没有得到类似结果。瓦杰(Tor D. Wager)等综合分析了在 1992 年到 2002 年 2 月进行的 65 项不同脑成像研究结果，得到的结论是：并没有充分的和一致的实验证据支持正负情绪加工的半球差异；分析认为情绪活动半球差异是很复杂的，并且具有很大的区域特殊性。

(3) 有证据表明激活维度和体验强度也并不是完全关联的。

### 12.2.6 非线性动态策略

情绪是一个随时间不断演变的多维度现象，用简单的线性模式去表达这样的多维系统，就必然会导致某些信息的损失，这妨碍了对于作为一个复杂过程的情绪的更全面的理解，于是特劳希(Tracy)等提出了情绪研究的非线性动态策略。

情绪不但包括许多互相联结的子系统，而且这些子系统的活动也是动态变化的。情绪是神经生理、外显表情和内在体验的综合过程，情绪的神经基础又包括了中枢神经系统、外

周神经系统和自主神经系统的活动。例如我们恐惧时,伴随着恐惧体验,我们同时会表现出恐惧的表情,以及一系列生理反应,如肌肉紧张、面色苍白、腹中空虚感、血液更多的流向四肢。所以在我们的情绪研究中,不能只从一个侧面研究情绪过程,也无法只凭借一种测量方法就能揭示情绪的科学规律,我们必须把这三个方面结合起来。

情绪是过程而不是状态,人们对于同一事件的情绪反应会随着时间而发生着动态变化,比如,对于一件新鲜事物,我们最初的反应是由于感到新奇而靠近,然而随着了解的深入,我们逐渐对其产生好感或者讨厌的情绪。谢莱(Schere)认为情绪是"各子系统的同步活动和动态联结"。特劳希等对由于看到蛇而诱发的恐惧情绪进行了血压、面部表情和恐惧体验的同步动态测量。测量结果是三者都呈现一定的波形,并最终回到稳态。这些反应的发生并不是彼此孤立的,各反应之间彼此联结——它们一起发生并互相影响。就像马歇尔(Marshall)和齐姆伯多(Zimbardo)曾经证明的那样,生理激活能够导致恐惧体验;情绪的行为表现(尤其是面部表情)又能导致生理唤醒和情绪体验。情绪是一个复杂的巨系统,只凭对各成分的了解无法对整个体系作出解释。

不同情绪之间彼此易感且相互作用,每次情绪都有可能受到在其前面发生的情绪的影响,同样也可能影响着下一次情绪。在心理生理研究中,早就引入了初始值原则(The Law of Initial Value,LIV),而对于情绪事件来说,一个重要的初始条件便是心境。心境本身也与先前的情绪事件密切相关。心境的主要功能是影响着特定情绪兴奋的阈值。一个冷刺激作用于皮肤会有2种不同情况,如果皮肤原本过热,该冷刺激就诱发快乐情绪,生理反应表现为皮肤毛细血管的收缩程度大;相反,如果原先皮肤就很冷,那么该冷刺激就成了负性刺激,而皮肤毛细血管的收缩程度较小;如果该冷刺激的温度高于初始的皮肤温度,那么反而会导致毛细血管舒张。同样,一个偶然事件(如在公交车上别人不小心踩了你一下)发生于快乐心境时,情绪系统不会被激活;然而如果该事件发生于愤怒心境时,则会激活情绪系统。这些都显示情绪的动态性依赖于系统的初始状态。

## 12.3 情绪加工

人类存在基本情绪,但是有关情绪加工的理论和研究主要还是针对焦虑和抑郁这两种情绪状态完成的,针对快乐的研究只有很少一部分,而对愤怒和厌恶的研究则几乎没有。一些情绪加工理论强调心境对情绪加工的作用,而另一些理论则关注人格因素对情绪加工的影响。然而,这两种理论之间实际上是存在重叠的。例如,我们可能想研究特质焦虑的影响因素。如果我们要做一个研究,那么那些具有高特质焦虑的被试很可能比低特质焦虑的被试处于更焦虑的心境状态。在这种情况下,我们很难分清人格和心境的作用。下面将介绍由鲍尔(G. H. Bower)、贝克(A. T. Beck)以及威廉斯(J. M. G. Williams)等提出的理论。

### 12.3.1 情绪语义网络理论

鲍尔与其助手所提出的网络理论的主要特点见图12.1,可以归纳为以下六个假设[93]。

(1) 情绪是语义网络中的单元或者结点,这些情绪结点与相关的观念、生理系统、事件、肌肉和表达模式等存在大量连接。

（2）情绪材料以命题或主张的形式储存于语义网络之中。

（3）思维通过激活语义网络中的结点而产生。

（4）结点可以被外部刺激或者内部刺激所激活。

（5）被激活的结点把激活扩散到与其相连的其他结点上。这个假设是相当关键的，因为这意味着一个情绪结点（如悲伤）的激活会引起语义网络中与情绪相关的结点或概念（如失落、绝望）的激活。

（6）“意识”是指网络中所有被激活结点的总激活量超过某一阈限值。

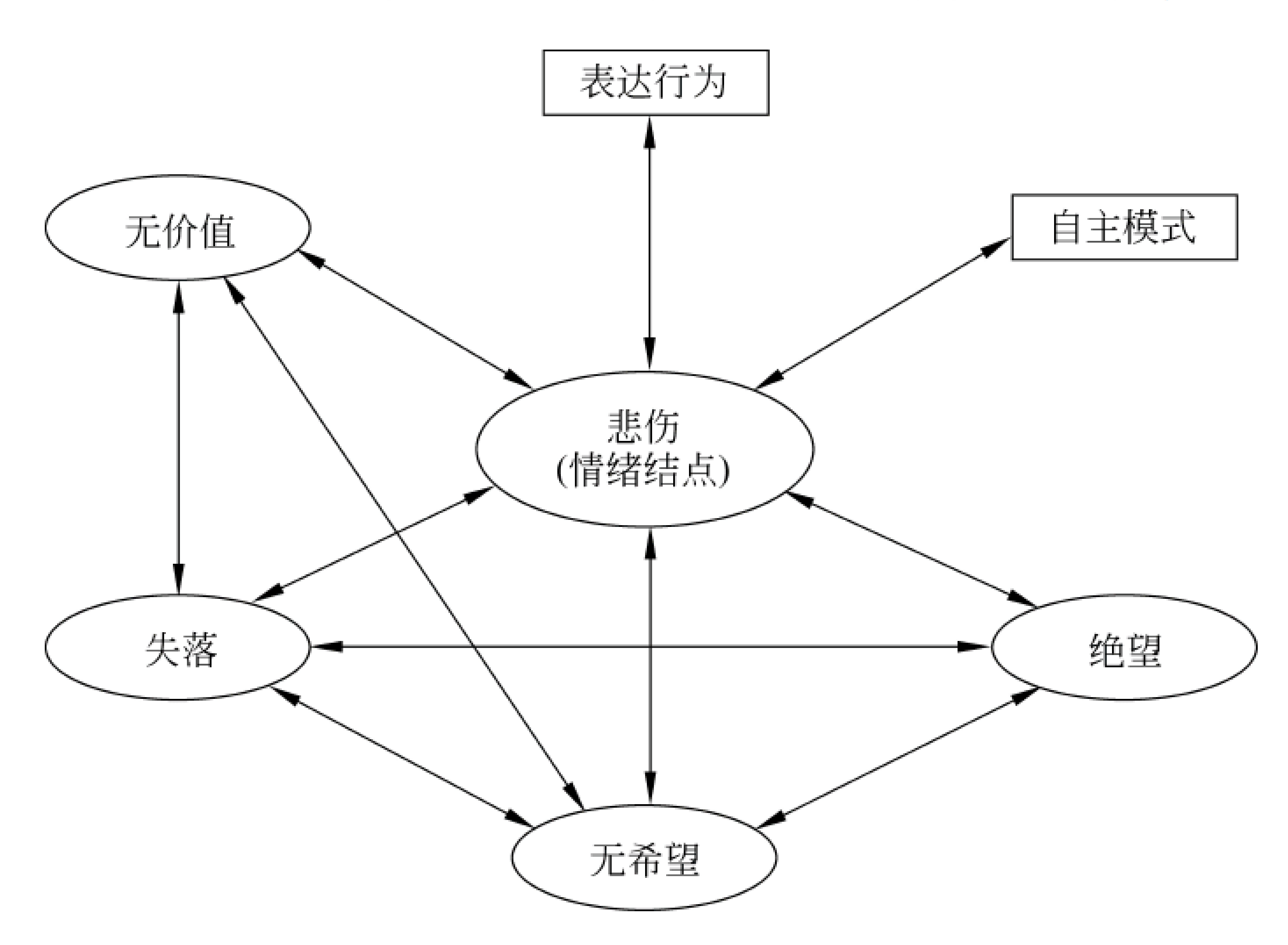



图 12.1　情绪语义网络理论

鲍尔的网络理论显得过于简单。这一理论把情绪或者心境以及认知概念都表征为语义网络中的结点。然而，心境和认知实际上差别很大。例如，心境在强度方面改变很慢，而认知往往是全或无的，常常是从一种认知加工迅速转变为另一种认知加工。

## 12.3.2　贝克的图式理论

贝克提出了一个图式理论[66]，核心内容是，某些人比其他一些人具有更高的易感素质（Vulnerability），易发展出抑郁或者焦虑障碍。这种易感素质取决于个体在早期生活经验中形成的某些图式或有组织的知识结构。贝克和克拉克的假设图式会影响大部分认知加工过程，如注意、知觉、学习和信息提取等。图式会引起加工偏向，即对图式一致性或情绪一致性信息的加工更受欢迎。如此一来，拥有焦虑相关图式的个体应该选择加工威胁性信息，而拥有抑郁相关图式的个体则选择加工负性情绪信息。虽然贝克和克拉克强调图式对加工偏向的作用，但他们认为只有当个体处于焦虑或者抑郁状态时，图式才会被激活并且会影响加工过程。

贝克的图式理论最初是为了给理解临床焦虑症和抑郁症提供一个理论框架。然而，该理论也可以应用于人格研究。某些个体拥有一些使他们表现出临床焦虑或抑郁症状的图式。这一观点是很有价值的。然而，要证明这种图式是引起焦虑障碍或者抑郁症的原因却是很困难的。这种方法存在如下一些缺陷。

(1) 图式的核心理论架构是模糊的,它常常不过是一种信念而已。

(2) 特定图式存在的证据常常是基于循环论证的。在焦虑症患者中,关于认知偏向的行为数据被用来推导图式的存在,然后这些图式又被用来解释所观察到的认知偏向。换句话说,通常不存在直接或独立的证据证明图式的存在。

### 12.3.3 威廉斯的情绪加工理论

威廉斯等关注的是焦虑和抑郁对情绪加工的影响[805]。他们是基于启动和精细加工之间的区别开始研究的。启动是一个自动加工过程。在启动条件下,一个刺激词激活长时记忆中该词的各个组成成分。而精细加工则是一个后期的策略加工过程,它涉及相关概念的激活。根据他们的理论,焦虑个体表现出对威胁刺激的初始启动效应,因此他们对威胁存在注意偏向。相反,抑郁个体表现出对威胁刺激的精细加工,所以他们对威胁刺激表现出记忆偏向,即发现他们提取威胁信息比提取中性信息要容易。

威廉斯等所做的一些主要预测是关于焦虑和抑郁对外显记忆和内隐记忆的影响作用。外显记忆是指有意识地回忆过去事件,这涉及精细加工。相反,内隐记忆不涉及有意识回忆,它主要依赖启动和自动加工过程。抑郁的个体应该表现出外显记忆偏向,喜欢以外显的记忆方式提取威胁性材料。而焦虑的个体则表现出内隐记忆偏向,喜欢以内隐的记忆方式提取威胁性信息。

研究结果更多地支持威廉斯等的理论范式,而支持鲍尔的网络理论和贝克的图式理论的证据则相对较少一些。例如,有很有力的证据证明焦虑与注意偏向有关,而证明抑郁与注意偏向相关的证据则弱得多。根据网络理论和图式理论,心境抑郁的个体对与心境状态一致的刺激的加工(和注意)应该更快,而且应该表现出对这类刺激材料的注意偏向。相反,威廉斯等认为抑郁个体不会给予威胁刺激过多的知觉加工,所以对这类刺激不会表现出注意偏向。威廉斯等的理论也可以较好地解释外显和内隐记忆偏向的研究结果。焦虑个体表现出对内隐记忆的偏向,抑郁的个体表现的是对外显记忆的偏向,这一预测得到一些研究的证实。

## 12.4 情感智能

智商测验不能全面衡量一个人的综合水准。对于智商高的人,他的其他智能并不一定成熟,其他的智能方面包括情感、艺术和体育等。换一句话说:智商高并不能保证他的未来就一定前途无量。过分强调先天的智慧,会把后天重要的能力培养部分忽略掉。

智力测验的缺陷主要是它太注重于语言和数理逻辑能力的重要性了。其实智能是多元的,它至少应该包括以下 7 种不同的智力:①言语智力;②数理逻辑智力;③空间智力;④音乐智力;⑤体能智力;⑥人际智力;⑦自知智力。这是嘉德纳初步对情感智能的概念的概述,为以后探讨情感智能作了有力的铺垫。

1990 年,萨拉维(P. Salovey)和梅耶尔(J. D. Mayer)正式提出了情感智能(Emotional Intelligence,EI)和情商(Emotional Quotient,EQ)的概念[634]。他们将情感智能定义为一种社会智能,包括监督自己和他人情绪的能力、区分自己和他人情绪的能力,以及运用情绪信

息去指导思维和行动的能力。情感智能包括以下 5 方面内容。

(1) 了解和表达自己情感的能力,真正知道自己确实感受的能力。

(2) 控制自己感情和延缓满足自己欲望的能力。

(3) 了解别人的情感以及对别的情感作出适当反应的能力。

(4) 能否以乐观态度对待挑战的能力。

(5) 处理人际关系的能力。

正如智商被用来反映传统意义上的智力一样,情商也被用来衡量一个人的情感商数的高低,主要是指人在情绪、意志、耐受挫折等方面的品质。如果说智商分数更多地是被用来预测一个人的学业成就,那么情商分数则被认为是用于预测一个人能否取得职业成功或生活成功的更有效的东西,它更好地反映了个体的社会适应性。

情商绝对无法用智商测验得知。为什么学校里成绩最优异的学生后来走入社会却难以成功。20 世纪 90 年代戈乐曼(D. Goleman)指出,智商的高低并不是决定一个人胜败的关键,而他本身具备的情商才是最为重要的因素。因为情商反映我们的自觉程度、冲动控制、坚持耐力、感染魅力、灵活程度和处事能力等方方面面。

一般来说,智商高者会被录用,但是情商高者往往更容易被提升。特别是在美国,许多大公司里藏龙卧虎,有无数顶尖大学毕业出来的高才生。然而这些人,由于一直很优秀,所以也容易过于独断高傲,难以与人相处。所以提升时,当然是那些平易近人、善解人意的部下会优先被考虑。这些人观察周围,观察人,把自己协调到合适的状态。

情商高的人能够控制自己的感情冲动,不求一时的痛快和满足;懂得如何激发自己不断努力;与人交往中善于理解别人的暗示,这样的人能了解人生遇到的荣辱成败。如果父母具备这些素质并能给予指导,孩子很容易具备这些素质。家长可以从以下几方面培养孩子的情感智能。

(1) 培养孩子正确的情绪反应,使孩子提早形成正确的情绪习惯。

(2) 学会准确表达自己的感觉。与人沟通往往因为不能准确表达各自的感觉和想法,而造成偏见和误会。

(3) 帮助孩子学会控制自己的欲望。家长可以通过生活中的事例让孩子明白,一个人想实现自己的愿望必须要经过不懈的努力,克服种种困难,否则是不可能的。

有关情感智能的研究可以参阅马特休斯(G. Matthews)等人的著作[461]。

## 12.5 情感计算

有关人类情感的研究,早在 19 世纪末就进行了,但是极少有人将“感情”和无生命的机器联系起来。让计算机具有情感能力是由美国麻省理工学院明斯基在 1985 年提出的,问题不在于智能机器能否有任何情感,而在于机器实现智能时怎么能够没有情感[498]。2006 年,明斯基发表专著《情感机器》[501]。他指出,情感是人类的一种特殊思维方式,提出了塑造智能机器的 6 大维度:意识、精神活动、常识、思维、智能、自我。

MIT 媒体实验室皮卡德(R. W. Picard)在 1997 年提出情感计算(Affective Computing)[569]。她指出,情感计算是关于情感、情感产生以及影响情感方面的计算。传统的人机交互,主要通过键盘、鼠标、屏幕等方式进行,只追求便利和准确,无法理解和适应人

的情绪或心境。而如果缺乏这种情感理解和表达能力,就很难指望计算机具有类似人一样的智能,也很难期望人机交互做到真正的和谐与自然。由于人类之间的沟通与交流是自然而富有感情的,因此,在人机交互的过程中,人们也很自然地期望计算机具有情感能力。情感计算就是要赋予计算机类似于人一样的观察、理解和生成各种情感特征的能力,最终使计算机像人一样能进行自然、亲切和生动的交互。

情感计算的目的是通过赋予计算机识别、理解、表达和适应人的情感的能力来建立和谐人机环境,并使计算机具有更高的、全面的智能。研究的重点就在于通过各种传感器获取由人的情感所引起的生理及行为特征信号,建立"情感模型",从而创建感知、识别和理解人类情感的能力,并能针对用户的情感做出智能、灵敏、友好反应的个人计算系统,缩短人机之间的距离,营造真正和谐的人机环境。情感计算主要研究内容如下。

(1) 情感机理的研究。情感机理的研究主要是情感状态判定及与生理和行为之间的关系。涉及到心理学、生理学、认知科学等,为情感计算提供理论基础。人类情感的研究已经是一个非常古老的话题,心理学家、生理学家已经在这方面做了大量的工作。任何一种情感状态都可能会伴随几种生理或行为特征的变化;而某些生理或行为特征也可能起因于数种情感状态。因此,确定情感状态与生理或行为特征之间的对应关系是情感计算理论的一个基本前提,这些对应关系目前还不十分明确,需要作进一步的探索和研究。

(2) 情感信号的获取。情感信号的获取研究主要是指各类有效传感器的研制,它是情感计算中极为重要的环节,没有有效的传感器,可以说就没有情感计算的研究,因为情感计算的所有研究都是基于传感器所获得的信号。各类传感器应具有如下基本特征:使用过程中不应影响用户(如重量、体积、耐压性等),应该经过医学检验对用户无伤害;数据的隐私性、安全性和可靠性;传感器价格低、易于制造等。MIT 媒体实验室的传感器研制走在了前面,已研制出多种传感器,如脉压传感器、皮肤电流传感器、汗液传感器及肌电流传感器等。皮肤电流传感器可实时测量皮肤的导电系数,通过导电系数的变化可测量用户的紧张程度。脉压传感器可时刻监测由心动变化而引起的脉压变化。汗液传感器是一条带状物,可通过其伸缩的变化时刻监测呼吸与汗液的关系。肌电流传感器可以测得肌肉运动时的弱电压值。

(3) 情感信号的分析、建模与识别。一旦由各类有效传感器获得了情感信号,下一步的任务就是将情感信号与情感机理相应方面的内容对应起来,这里要对所获得的信号进行建模和识别。由于情感状态是一个隐含在多个生理和行为特征之中的不可直接观测的量,不易建模,部分可采用诸如隐马尔科夫模型、贝叶斯网络模式等数学模型。MIT 媒体实验室给出了一个隐马尔科夫模型,可根据人类情感概率的变化推断得出相应的情感走向。研究如何度量人工情感的深度和强度,研究定性和定量的情感度量的理论模型、指标体系、计算方法、测量技术。

(4) 情感理解。通过对情感的获取、分析与识别,计算机便可了解其所处的情感状态。情感计算的最终目的是使计算机在了解用户情感状态的基础上,作出适当反应,去适应用户情感的不断变化。因此,这部分主要研究如何根据情感信息的识别结果,对用户的情感变化做出最适宜的反应。在情感理解的模型建立和应用中,应注意以下事项:情感信号的跟踪应该是实时的和保持一定时间记录的;情感的表达是根据当前情感状态、实时的;情感模型是针对于个人生活的,并可在特定状态下进行编辑;情感模型具有自适应性;通过理解

情况反馈调节识别模式。

(5) 情感表达。前面的研究是从生理或行为特征来推断情感状态。情感表达则是研究其反过程,即给定某一情感状态,研究如何使这一情感状态在一种或几种生理或行为特征中体现出来,例如如何在语音合成和面部表情合成中得以体现,使机器具有情感,能够与用户进行情感交流。情感的表达提供了情感交互和交流的可能,对于单个用户来讲,情感的交流主要包括人与人、人与机器、人与自然和人类自己的交互、交流。

(6) 情感生成。在情感表达基础上,进一步研究如何在计算机或机器人中,模拟或生成情感模式,开发虚拟或实体的情感机器人或具有人工情感的计算机及其应用系统的机器情感生成理论、方法和技术。

到目前为止,有关研究已经在脸部表情、姿态分析、语音的情感识别和表达方面获得了一定的进展。

### 1. 脸部表情

在生活中,人们很难保持一种僵硬的脸部表情,通过脸部表情来体现情感是人们常用的较自然的表现方式,其情感表现区域主要包括嘴、脸颊、眼睛、眉毛和前额等。人在表达情感时,只稍许改变一下面部的局部特征(譬如皱一下眉毛),便能反映一种心态。在 1972 年,著名的学者埃克曼提出了脸部情感的表达方法(脸部运动编码系统 FACS)。通过不同编码和运动单元的组合,即可在脸部形成复杂的表情变化,譬如幸福、愤怒、悲伤等。该成果已经被大多数研究人员所接受,并被应用于人脸表情的自动识别与合成(图 12.2)。

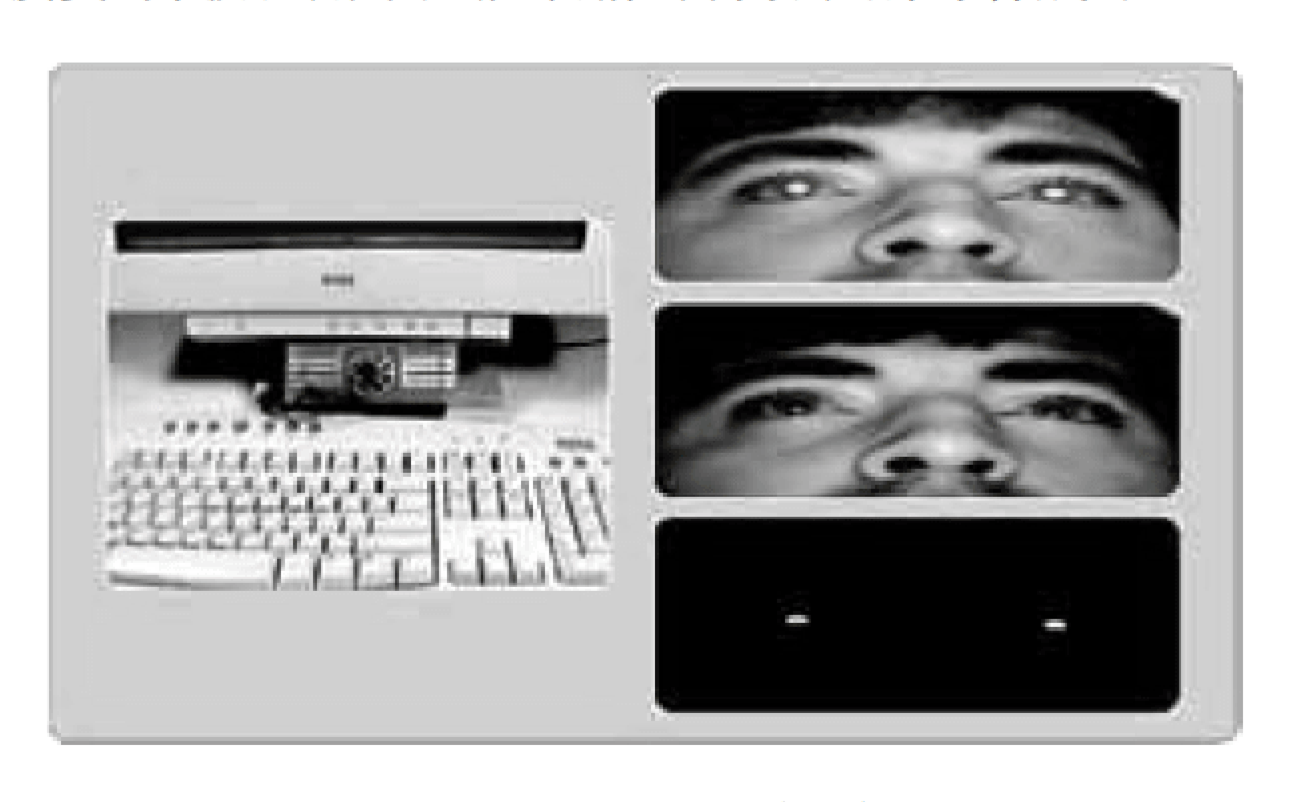


图 12.2　表情的识别与合成

随着计算机技术的飞速发展,为了满足通信的需要,人们进一步将人脸识别和合成的工作融入到通信编码中。最典型的便是 MPEG4 V2 视觉标准,其中定义了 3 个重要的参数集:人脸定义参数、人脸内插变换和人脸动画参数。表情参数中具体数值的大小代表人激动的程度,可以组合多种表情以模拟混合表情。

在目前的人脸表情处理技术中,多侧重于对三维图像的更加细致的描述和建模。通常采用复杂的纹理和较细致的图形变换算法,达到生动的情感表达效果。在此基础上,不同的算法形成了不同水平的应用系统。

### 2. 姿态变化

人的姿态一般伴随着交互过程而发生变化,它们表达着一些信息。例如手势的加强通常反映一种强调的心态,身体某一部位不停地摆动,则通常具有情绪紧张的倾向。相对于语音和人脸表情变化来说,姿态变化的规律性较难获取,但由于人的姿态变化会使表述更加生动,因而人们依然对其表示了强烈的关注。科学家针对肢体运动,专门设计了一系列运动和身体信息捕获设备,例如运动捕获仪、数据手套、智能座椅等。国外一些著名的大学和跨国公司,例如麻省理工学院、IBM 等则在这些设备的基础上构筑了智能空间。同时也有人将智能座椅应用于汽车的驾座上,用于动态监测驾驶人员的情绪状态,并提出适时警告。意大利的一些科学家还通过一系列的姿态分析,对办公室的工作人员进行情感自动分析,设计出更舒适的办公环境。

### 3. 语音理解

在人类的交互过程中,语音是人们最直接的交流通道,人们通过语音能够明显地感受到对方的情绪变化,例如通过特殊的语气词、语调发生变化等等。在人们通电话时,虽然彼此看不到,但能从语气中感觉到对方的情绪变化。例如同样一句话"你真行",在运用不同语气时,可以使之成为一句赞赏的话,也可以使之成为讽刺或妒忌的话。

目前,国际上对情感语音的研究主要侧重于情感的声学特征的分析这一方面。一般来说,语音中的情感特征往往通过语音韵律的变化表现出来。例如,当一个人发怒的时候,讲话的速率会变快,音量会变大,音调会变高等,同时一些音素特征(共振峰、声道截面函数等)也能反映情感的变化。中国科学院自动化研究所模式识别国家重点实验室的专家们针对语言中的焦点现象,首先提出了情感焦点生成模型。这为语音合成中情感状态的自动预测提供了依据,结合高质量的声学模型,使得情感语音合成和识别率先达到了实际应用水平。

### 4. 多模态的情感计算

虽然人脸、姿态和语音等均能独立地表示一定的情感,但人在相互交流的过程中却总是通过上面信息的综合表现来进行的。所以,唯有实现多通道的人机界面,才是人与计算机最为自然的交互方式,它集自然语言、语音、手语、人脸、唇读、头势、体势等多种交流通道为一体,并对这些通道信息进行编码、压缩、集成和融合,集中处理图像、音频、视频、文本等多媒体信息。目前,多模态技术本身也正在成为人机交互的研究热点,而情感计算融合多模态处理技术,则可以实现情感的多特征融合,能够有力地提高情感计算的研究深度,并促使出现高质量、更和谐的人机交互系统。

### 5. 情感计算与个性化服务

随着情感计算研究的进一步深入,人们已经不仅仅满足于将其应用在简单的人机交互平台中,而要拓展到广泛的界面设计、心理分析、行为调查等各个方面,以提高服务的质量,并增加服务的个性化内容。在此基础上,有人开始专门进行情感主体(Affective Agent)的研究,以期通过情感交互的行为模式,构筑一个能进行情感识别和生成的类生命体,并以这

个模型代替传统计算中的有些应用模型中(例如电脑游戏的角色等),使电脑和应用程序更加鲜活起来,使之能够产生类似于人的一些行为或思维活动。这一研究还将从侧面上对人工智能的整体研究产生较大的推动作用。

### 6. 情感理解模型

情感状态的识别和理解,则是赋予计算机理解情感并做出恰如其分反应的关键步骤。这个步骤通常包括从人的情感信息中提取用于识别的特征,例如从一张笑脸中辨别出眉毛等,接着让计算机学习这些特征以便日后能够准确地识别其情感。

为了使计算机更好地完成情感识别任务,科学家已经对人类的情感状态进行了合理而清晰的分类,提出了几类基本情感。目前,在情感识别和理解的方法上运用了模式识别、人工智能、语音和图像技术的大量研究成果。例如:在情感语音的声学分析的基础上,运用线性统计方法和神经网络模型,实现了基于语音的情感识别原型;通过对面部运动区域进行编码,采用 HMM 等不同模型,建立了面部情感特征的识别方法;通过对人姿态和运动的分析,探索肢体运动的情感类别等。

不过,受到情感信息的捕获技术的影响,并缺乏大规模的情感数据资源,有关多特征融合的情感理解模型的研究还有待深入。随着未来的技术进展,还将提出更有效的机器学习机制。

情感计算与智能交互技术试图在人和计算机之间建立精确的自然交互方式,将会是计算技术向人类社会全面渗透的重要手段。未来随着技术的不断突破,情感计算的应用势在必行,其对未来日常生活的影响将是方方面面的,目前我们可以预见的有:

(1) 情感计算将有效地改变过去计算机呆板的交互服务,提高人机交互的亲切性和准确性。一个拥有情感能力的计算机,能够对人类的情感进行获取、分类、识别和响应,进而帮助使用者获得高效而又亲切的感觉,并有效减轻人们使用电脑的挫败感,甚至帮助人们便于理解自己和他人的情感世界。

(2) 情感计算还能帮助我们增加使用设备的安全性(例如当采用此类技术的系统探测到司机精力不集中时可以及时改变车的状态和反应)、使经验人性化、使计算机作为媒介进行学习的功能达到最佳化,并从我们身上收集反馈信息。例如,一个研究项目在汽车中用电脑来测量驾车者感受到的压力水平,以帮助解决所谓驾驶者的"道路狂暴症"问题。

(3) 情感计算和相关研究还能够给涉及电子商务领域的企业带来实惠。已经有研究显示,不同的图像可以唤起人类不同的情感。例如,蛇、蜘蛛和枪的图片能引起恐惧,而有大量美元现金和金块的图片则可以使人产生非常强烈的积极反应。如果购物网站和股票交易网站在设计时研究和考虑这些因素的意义,将对客流量的上升产生非常积极的影响。

(4) 在信息家电和智能仪器中,增加自动感知人们的情绪状态的功能,可以提供更好的服务。在信息检索应用中,通过情感分析的概念解析功能,可以提高智能信息检索的精度和效率。在远程教育平台中,情感计算技术的应用能增加教学效果。

(5) 利用多模式的情感交互技术,可以构筑更贴近人们生活的智能空间或虚拟场景等。情感计算还能应用在机器人、智能玩具、游戏等相关产业中,以构筑更加拟人化的风格和更加逼真的场景。

## 12.6　情感与认知

对刺激进行认知加工是情感反应发生改变的必要前提吗？这个问题在理论上具有重要意义。如果人类对一切刺激的情感反应都取决于认知加工，那么所有情绪理论都必须带有显著的认知特色。相反，如果认知加工在情感反应的发展过程中不是必需的，那么我们就没有必要把认知范式运用到情绪研究中去。

### 12.6.1　情感优先假说

扎琼克(R B Zajonc)认为对刺激的情感评价可以不依赖认知加工而独立进行[823]。按照他的观点，情感与认知是分开的，是部分独立的两个系统[824]。

根据情感优先假说，加工刺激的简单情感属性要比加工更高级的认知属性在速度上要快得多[514]。墨菲(S T Murphy)和扎琼克通过一系列实验研究对这一假说提供了一些支持。实验范式是，在启动刺激呈现 4ms 或 1s 后，紧接着呈现第二个刺激。在一个研究中，启动刺激是一些高兴或生气的面孔。同时，实验者还设计了一个无启动刺激的控制条件。启动刺激之后的第二个刺激是一个象形汉字。刺激呈现完毕后，被试需要对这些汉字的偏好度进行等级评定。当启动刺激的呈现时间为 4ms 时，被试对汉字的偏好度受到情感性启动刺激的影响，但是当启动刺激的呈现时间为 1s 时，这些启动刺激就没有影响。这可能是因为在后一种呈现条件下，被试认识到自己的情感反应源自启动刺激，因此情感反应没有影响他们对第二个刺激的评价。墨菲和扎琼克所获得的一系列结果支持下述结论。

(1) 情感加工有时候会比认知加工更快。

(2) 对一个刺激的早期情感加工与后期认知加工之间存在很大差异。

### 12.6.2　认知评价观点

拉扎勒斯(R S Lazarus)认为某些认知加工是情感反应发生的必要先决条件，认为认知评价对情绪体验起关键作用[399]。认知评价可以细分为三种更为具体的评价形式。

(1) 初级评价：把周围情境看成是积极的、有压力的或者是与幸福无关的。

(2) 次级评价：根据个体可以利用的情景资源来进行评价。

(3) 重新评价：刺激情境及相应的应付策略得到监控，必要时还要修改初级评价和次级评价。

拉扎勒斯通过一系列研究发现，各种心生理指标揭示否定式和理智式的指导语确实都减少了被试所承受的压力。因此，在个体面临应激事件时操控被试的认知评价会对他们的生理应急反应造成重要影响。认知评价总是发生在情感反应之前，不过这种评价可能并不一定发生在意识水平之上。然而，前意识的认知加工决定情感反应这一观点通常只不过是一个个人信仰问题。不过有关意识下知觉的研究文献表明前意识认知加工确实存在，而且具有重要意义。

### 12.6.3 图式命题联想和类比表征系统

普尔(M Power)和达格莱希(T Dalgleish)提出一个图式命题联想和类比表征系统(Schematic Propositional Associative and Analogical Representational System,SPAARS)[592],具体结构见图 12.3。模型中的各种成分阐述如下。

(1) 类比系统:这一系统涉及对环境刺激进行基本的感觉信息加工。

(2) 命题系统:这一系统不涉及情绪因素,它只包含对外界和自我的信息。

(3) 图式系统:在这一系统中,来自命题系统的事实与来自个体近期目标的有关信息结合在一起,生成一个针对情境的内部模型,从而引起情绪反应。

(4) 联想系统:如果同一个事件在图式水平以相同的方式不断地被加工,那么一个联想表征就会形成。当将来遇上同一事件时,个体的相关情绪就会被自动引发出来。

普尔和达格莱希认为,情绪的产生主要有两种方式。首先,当图式系统参与信息加工时,情绪完全是认知加工的结果。第二,当联想系统参与信息加工时,情绪可以在不涉及有意识加工的情况下自动发生。

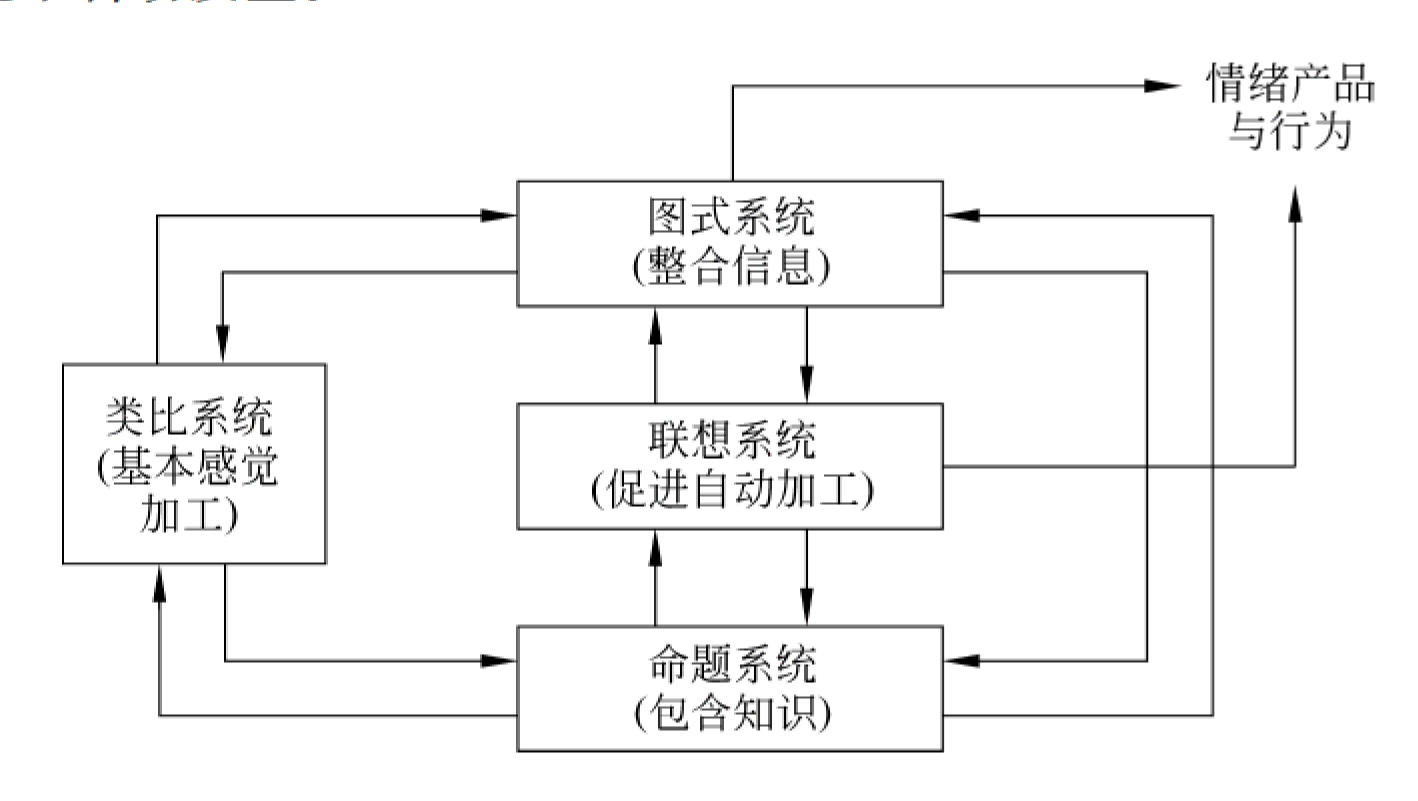



图 12.3 图式命题联想和类比表征系统 SPAARS

## 12.7 情绪的脑机制

情绪是人脑的高级功能,保证着有机体的生存和适应,对个体的学习、记忆、决策有着重要的影响。情绪的脑机制——大脑回路,包括前额皮层、杏仁核、海马、前部扣带回、腹侧纹状体等。前额皮层中的不对称性与趋近和退缩系统有关,左前额皮层与趋近系统和积极感情有关,右前额皮层与消极感情和退缩有关。杏仁核易被消极的感情刺激所激活,尤其是恐惧。海马在情绪的背景调节中起着重要作用。前额皮层和杏仁核激活不对称性的个体差异是情绪个体差异的生理基础。

情绪由大脑中的一个回路所控制,包括前额皮层、杏仁核、海马、前部扣带回、腹侧纹状体等。它们整合加工情绪信息,产生情绪行为。许多文献表明,有两个基本的情绪和动机系统或者积极和消极感情形式,分别是趋近和退缩。1999 年,戴维森(R J Davidson)等人把趋近系统描述为促进欲求行为和产生特定的与趋近有关的积极感情类型[173],如愉快、兴趣等。退缩系统有利于有机体从厌恶刺激源撤退或者组织对威胁线索的适当反应,产生与撤

退有关的消极情绪,如厌恶和恐惧等。各种证据表明,趋近和退缩系统是由部分独立的回路执行的。

### 1. 前额皮层

灵长类动物的前额皮层可分为 3 个子分区:背侧 PFC(DLPFC)、腹内侧 PFC(vmPFC)、眶额皮层(OFC)。前额皮层的各个部分与情绪有关。左前额皮层与积极感情有关,右前额皮层与消极感情有关。

在已有研究的基础上,米勒(E K Miller)和科恩提出了一个综合的前额机能理论,认为前额皮层 PFC 维持对目标的表征和达到目标的方法[495]。腹内侧前额皮层与对未来积极和消极感情后果的期待有关。贝卡拉(A Bechara)等人于 1994 年报告腹内侧前额皮层两侧损伤的病人在期待未来的积极和消极后果中有困难[65]。这样的病人与控制组相比,在期待冒险选择中,表现出皮肤电活动水平的降低。

### 2. 杏仁核

杏仁核对知觉、产生消极感情和联想厌恶学习很重要。对恐惧面部表情的反应中杏仁核激活。许多研究报告在厌恶条件作用的早期阶段杏仁核激活。对几个诱发消极感情实验程序的反应中也可观察到杏仁核激活,包括厌恶嗅觉线索和厌恶味觉刺激等。

双侧杏仁核损伤病人对恐惧和愤怒声音的识别有困难,表明这一缺陷并不限于面部表情。杏仁核损伤病人对厌恶刺激无反应。总之,研究结果表明,双侧杏仁核受破坏的病人加工消极情绪任务的能力被损害,表明杏仁核对识别威胁或危险线索是重要的。

利铎克斯(J E LeDoux)关注焦虑这一情绪反应,他强调杏仁核的作用,把它看作是大脑的"情绪计算机",负责计算出刺激的情绪价值[404]。根据利铎克斯的观点,情绪刺激的感觉信息是从丘脑同时传送到杏仁核和大脑皮质的。在此基础上,利铎克斯提出焦虑存在两条不同的情绪回路。

(1) 一条是"丘脑→大脑皮质→杏仁核"这一慢回路,它负责对感觉信息进行详细分析。

(2) 另一条是"丘脑→杏仁核"这一快回路,它负责对刺激的简单特征(如刺激强度)进行加工。这条回路无须经过大脑皮质。

来自大脑新皮质的信号对杏仁核的激活与情绪加工发生在认知加工之后这一传统观点是吻合的,而来自丘脑的信号对杏仁核的激活与情感优先假说是一致的,即情绪加工可以发生在前意识水平而且是发生在认知加工之前的。丘脑→杏仁核回路使我们能够对危险情景做出快速反应,因而这条回路在保障我们的生存方面很有价值。相反,皮质回路使我们可以详细评价情境的情绪意义,让我们能以最佳方式对情境做出反应。

### 3. 海马和前扣带回

海马在情绪中的作用近年来才开始研究。海马是大脑中有很高葡萄糖皮质激素类受体密度的部位,在情绪调节中很重要。戴维森等人提出海马在情绪行为的背景调节中起关键作用。如果海马损害则个体正常背景的调节作用受到损害,因而在不适当的背景中表现出情绪行为。研究发现,对赢钱和输钱左和右杏仁核有不同的激活,左侧杏仁核对赢更多的钱显示激活的提高,而右侧杏仁核对输钱显示激活的提高。

神经成像方法的研究表明前扣带回在情绪反应中激活。对情绪单词的 Stroop 任务(一个刺激的两个不同维度发生相互干扰的现象)的反应中,观察到背侧前扣带回激活。

## 4. 腹侧纹状体

PET 研究中观察到,在图片诱发感情期间,听神经核的腹侧纹状体区域被激活。发现被试在看愉快的录相游戏时,这一区域中的多巴胺水平提高。

情绪是人脑的高级功能,是人类生存适应的第一心理工具。它具有组织、调节和动机的功能。是个性的核心内容,也是控制心理病理的关键成分。因此对情绪发生、发展脑机制规律的揭示,有利于促进个体智力的发展、身心的健康,使个体形成良好的个性。

# >>第 13 章
# 意识

意识的起源与本质是最重大的科学问题之一。在智能科学中，意识问题具有特别的挑战意义。存在如何决定意识，客观世界如何反映到主观世界中去，既是哲学研究的主题，也是当代自然科学研究的重要课题。意识涉及知觉、注意、记忆、表征、思维、语言等高级认知过程，其核心是觉知(Awareness)。近年来，由于认知科学、神经科学和计算机科学的发展，特别是新的无损伤性实验技术的出现，意识的研究再度被提到日程上来，并且开始成为众多学科共同研究的热点。在 21 世纪，意识问题将是智能科学力图攻克的难题之一。

## 13.1 概述

意识(Consciousness)是一种复杂的生物现象，哲学家、医学家、心理学家对于意识的概念各不相同，迄今尚无定论。当代著名思想家丹尼特(D C Dennett)认为[181]："人类的意识大概是最后一个难解的谜……对意识，我们至今如坠五里云雾中，时至今日，意识是唯一常常使最睿智的思想家张口结舌、思维混乱的论题。"

意识的哲学概念是高度完善、高度有组织的特殊物质——人脑的机能，是人所特有的对客观现实的反映。意识也作为思维的同义词，但意识的范围较广，包括认识的感性和理性阶段，而思维则仅指认识的理性阶段。辩证唯物主义认为意识是物质高度发展的产物，是存在的反映，又对存在起着巨大的能动作用。

医学上，不同学科对意识的认识也略有差异。在临床医学领域，意识的概念是指病人对周围环境及自身的认识和反应能力，分为意识清楚、意识模糊、昏睡、昏迷等不同的意识水平；在精神医学中，意识又有自我意识和环境意识的分别。意识障碍表现为意识浑浊、嗜睡、昏睡、昏迷、谵妄、朦胧状态、梦样状态和意识模糊。

心理学对意识的观点是对外部环境和自身心理活动，例如感觉、知觉、注意、记忆、思想等客观事物的觉知或体验。进化生物学家、理论神经科学家威廉·卡尔文(William H Calvin)在《大脑如何思维》一书中列出了一些意识的定义[125]。

从智能科学的角度，意识是一种主观体验，是对外部世界、自己的身体及心理过程体验的整合。意识是一种大脑本身具有的"本能"或"功能"，是一种"状态"，是多个脑结构对于多种生物的"整合"。广义的意识是高等生物与低等生物都具有的一种生命现象。随着生物的进化，进行意识加工的器官也在不断进化。人类进行意识活动的器官主要是脑。为了揭示意识的科学规律，构建意识的脑模型，不仅需要研究有意识的认知过程，而且需要研究无意识的认知过程，即脑的自动信息加工过程，以及这两种过程在脑内的相互转化机制。意识研究是认知神经科学不可缺少的内容，意识及其脑机制的研究是自然科学的重要内容。哲学

所涉及的是意识的起源和意识存在的真实性等问题,意识的智能科学研究的核心问题是意识产生的脑机制——物质的运动如何变成意识的。

历史上最早使用意识这个词的是培根(Francis Bacon)。他的定义是意识就是一个人对自己思想里发生了什么的认识。所以,意识问题一直是哲学家研究的领域。德国心理学家冯特(Wundt)于 1879 年建立了第一个心理学实验室,明确提出心理学主要是研究意识的科学,以生理学方法研究意识,报告在静坐、工作和睡眠条件下的意识状态。从此心理学以一门实验科学的身份进入了一个新的历史时期,一系列心理现象的研究都得到了迅速发展,但是意识的研究因缺少非意识的直接客观指标而进展迟缓。1902 年加米斯(James)提出意识流的概念,指出意识就像流水一样波浪起伏,渊源不断。Freud 认为人的感觉和行为受非意识需要、愿望和冲突的影响。根据 Freud 的观点,意识流具有深度,意识与非意识加工有不同的认识水平。它不是全或无的现象。但是,由于当时科学不够发达,用内省法进行,缺乏客观指标,只能停留在描述性初级水平上而无法前进。但是自从华生宣告心理学是一门行为科学之日起,意识问题被打入冷宫。所以有很长一段时间,神经科学因其太复杂而不敢问津,心理学又不愿染指被人遗忘的科学。

在 20 世纪 50～60 年代,科学家们通过解剖学、生理学实验来理解意识状态的神经生理学基础。例如,1949 年莫罗兹(Moruzziz)与马戈恩(Magoun)发现了觉知的网状激活系统;1953 年阿塞林斯基(Aserinsky)与克雷特曼(Kleitman)观察了快速眼动睡眠的意识状态;20 世纪 60～70 年代,进行了对割裂脑病人的研究,支持在大脑两半球中存在独立的意识系统。上述研究结果开创并奠定了意识的认知神经科学研究基础。

现代认知心理学始于 20 世纪 60 年代,对于认知心理学家来说,阐明客观意识的神经机制始终是一个长期的挑战。迄今关于意识客观体验与神经活动关系的直接研究还非常少见。近年,随着科学技术的突飞猛进,利用现代电生理技术(脑电图 EEG,事件相关电位 ERP)和放射影像技术(正电子断层扫描 PET,功能磁共振成像 FMRI),意识研究已迅速成为生命科学和智能科学的新生热点。

关于意识脑机制的研究虽然非常复杂,任务艰巨,但意义重大,已引起了全世界认知科学、神经生理、神经成像和神经生物化学等神经科学、社会科学以及计算机科学诸多领域学者们的极大兴趣。1997 年,国际意识科学研究学会(Association for the Scientific Study of Consciousness, ASSC)成立,连续召开意识问题国际学术会议。会议主题分别是:内隐认知与意识的关系(1997 年);意识的神经相关性(1998 年);意识与自我知觉和自我表征(1999 年);意识的联合(2000 年);意识的内容:知觉、注意和现象(2001 年);意识和语言(2002 年);意识的模型和机理(2003 年);意识研究中的经验和理论问题(2004 年)。

研究意识问题的科学家所持的观点是多种多样的。从人的认识能力最终是否有可能解决意识问题考虑,有神秘主义和还原论(Reductionism)之分。持神秘主义观点的人认为我们永远无法理解意识。例如,当代著名哲学家 Fodor 参加第一次 Towards to science of consciousness 会时公开怀疑:任何一种物理系统怎么会具有意识状态呢?在意识问题研究中十分活跃的美国哲学家查尔莫斯(D J Chalmers)认为,意识应当分为“容易问题”(Easy Problem)和“艰难问题”(Hard Problem)[131],他对意识问题的总看法是:“没有什么严谨的物理理论(量子机制或神经机制)可以理解意识问题。”

克里克(F Crick)在《惊人的假设》[164]一书中公开申明对意识问题的看法是还原论的,

他和他的年轻的追随者 Koch 在许多文章中陈述这一观点。他们把这个复杂的意识问题"还原"成神经细胞及其相关分子的集体行为。美国著名的计算神经科学家索诺斯基(Terrence J Sejnowski)和美国哲学家丹尼特(Daniel C Dennett)等人所持观点,大体上与克里克相同。

在研究意识问题时,从所持的哲学观点考虑,历来就有两种相反的观点:一种是一元论,认为精神(包括意识)是由物质(脑)产生的,是可以从脑来研究和解释精神现象的;另一种是二元论,认为精神世界独立于人体(人脑)。二者之间没有直接的联系。笛卡儿(René Descartes)是典型的二元论者,他认为每个人都有一个躯体和一个心灵(Mind)。人的躯体和心灵通常是维系在一起的,但心灵的活动不受机械规律的约束。躯体死亡后,心灵将继续存在,并且还发挥作用。一个人的心灵所进行的种种活动是无法被他人察知的,因此只有我才能直接知觉我个人的内心的状态和过程。如果把身体比拟为"机器",按照物理规律运行,那么,心灵就是"机器中的灵魂"。笛卡儿是伟大的数学家,所以他有正视现实的一面,在科学上明确地提出"人是机器"的论断,但他受古代哲学思想和当代社会环境的影响较深,所以他把脑的产物(精神)看成是与人体截然分开的东西。

在当代从事自然科学研究的科学家中间,有不少相信二元论的。诺贝尔奖获得者埃克尔斯(John Carew Eccles),热衷于意识问题的研究。他本人是神经科学家,研究神经细胞的突触结构和功能取得重大成果。他不讳言他的意识观是二元论的。他本人以及与人合作的关于脑的功能方面的著作有七本之多,他在与哲学家 Popper 的著作中提出"三个世界"的理论,其中第一世界是物理世界,包括脑的结构和功能,第二世界是所有主观精神和经验,第三世界是社会、科学和文化活动。他在后期的著作中,根据神经系统的结构和功能,提出"树突子"的假设,树突子是神经系统的基本结构和功能单元,由 100 个左右顶部树突构成。估计在人脑中有 40 万个树突子。他进而又提出"心理子"的假设,第二世界的心理子与第一世界的树突子相对应。由于树突中的微结构与量子尺度相近,所以量子物理有可能用于意识问题。

意识问题的研究需要靠人来进行,特别需要用人脑去研究,这就涉及人脑能否理解人脑的问题,因此有人说,用手把自己头发拉起是不可能做到脱离地球的。实际上意识问题上的一元论者和二元论者之间,可知论与不可知论之间,唯物论与唯心论之间,均不是界限截然分明的。

## 13.2 意识的基本要素和特性

法伯(I B Farber)和丘奇朗德(P S Churchland)在其《意识与神经科学,哲学与理论问题》一文中,从三个层次讨论了意识概念[217]。第一个层次是意识觉知,包括感觉觉知(指通过感觉通道对外部刺激的觉知)、概括性觉知(是指与任一感觉通道都不相连的对身体内部状态的觉察,如疲劳、眩晕、焦虑、舒服、饥饿等)、元认知觉知(是指能觉察到自己认知范围内的所有事物,包括当前的和过去的思维活动)和有意识回忆(能觉察到过去发生的事情)四种。这里所说的能觉察到某事物的标志,即能用言语报告该事物。这样既便于检测,也可以把不能说话的动物排除在外。第二个层次是高级能力,即不仅能被动地感知和觉知信息,还具有能动作用或控制等高级功能,这些功能包括注意、推理和自我控制(如理性或道德观念

对生理冲动的抑制作用)。第三个层次是意识状态,可理解为一个人正在进行的心理活动,包括意识概念中最常识性的也是最困难的环节,这种状态可以分为不同的层次,如有意识与非意识、综合性调节、粗略的感觉等。法伯的前两个层次对意识给出的定义是颇有启发性的,但第三层次却缺乏实质性内容。

1977 年,奥恩斯泰(R E Ornstein)提出意识存在的两种模式——主动-言语-理性模式(Active-verbal-rational Mode)与感知-空间-直觉-整体模式(Receptive-spatial-intuitive-holistic Mode)[547],分别简称为主动模式和感知模式。他认为两种模式分别被一侧大脑半球所控制,对主动模式的评价是自动进行的,人类限制了觉知的自动化以阻挡与其生存能力无直接相关的经验、事件和刺激。当人们需要加强正在进行的归纳与判断时,通过感知模式增加了正常的觉知。根据奥恩斯泰(R E Ornstein)的观点,静坐、生物反馈、催眠,甚至试验某些特异性药物也能有助于学习使用感知模式来平衡主动模式。智力活动是主动发生的,具有左半球优势,而直觉行为是感受性的,为右半球优势的。两种模式的整合构成了人类高级功能的基础。

意识功能是由哪些要素构成的?关于这一问题,克里克认为意识至少包括两个基本功能部件[164],一是注意,二是短时记忆。注意一直是意识的主要功能,这已为大家所公认。巴尔斯(B J Baars)的"剧场"隐喻中,把意识比喻为一个舞台,不同的场景轮流上场。平台上的聚光灯可比喻为注意机制,这是一个流行的比喻。克里克也认可这个比喻。没有记忆的人肯定没有"自我意识"。没有记忆的人或机器,看过即忘,或听过即忘,也不能妄谈意识。但记忆的时间长短可以讨论,长时记忆固然重要,克里克认为短时记忆更显必要。

美国哲学家与心理学家詹姆士(William James)认为意识有如下特点。

(1) 意识是个人的,不能与他人共享。

(2) 意识是永远变化的,不会长久停留在某一种状态。

(3) 意识是连续的,一个内容包含着另一个内容。

(4) 意识是有选择性的。

总之,詹姆士认为意识不是一个东西,而是一种过程,或一种"流",是一种可以在几分之一秒内变化的过程。这种"意识流"概念,很生动地刻画他关于意识的一些特性,这一概念在心理学中受到重视。

埃德尔曼(G M Edelman)强调意识的整合性和分化性[203]。依据脑的生理病理和解剖学上的事实,埃德尔曼认为丘脑-皮质系统在意识的产生方面起关键作用。

美国心脑问题的哲学家丘奇兰德(P S Churchalnd)为意识问题列出一张特性表[146]。

- 与工作记忆有关。
- 不依赖感觉输入,即我们能思考并不存在的东西和想像非真实的东西。
- 表现出可驾驭的注意力。
- 有能力对复杂或模棱两可的资料作出各种解释。
- 在深睡时消失。
- 在梦中重新出现。
- 在单次统一的经验中能包容若干感觉模态的内容。

2012 年,巴尔斯和埃德尔曼在文章[49]中阐述他们关于意识的自然观,列出了意识状态的 17 个特性。

(1) 意识状态的 EEG 标记。

脑的电生理活动呈现不规则、低幅度和快速的电活动,频率从 0.5Hz 到 400Hz。意识 EEG 看起来与无意识状态(类同沉睡情况)显著不同,癫痫患者和全身麻醉的意识状态呈现规则、高幅度、慢变化的电压。

(2) 大脑和视丘。

意识取决于是视丘的复杂性,开启和关闭通过脑干调制,并且与脑皮层下区域没有交互作用,不直接支持意识经验。

(3) 广泛的大脑活动。

可报告意识事件与广泛的具体脑活动内容有关。无意识的刺激只唤起局部的脑活动。意识瞬间也对外边专注意识内容引发广泛的影响,表现为隐性学习、情景记忆、生物反馈训练等。

(4) 大范围的可报告内容。

意识有特别广泛的不同内容——各种感觉的知觉、内生的形象化描述、感情感觉、内部语言、概念、有关行动的想法和像熟悉的感觉那样的外部经验。

(5) 信息性。

当信号变得多余时意识可以消失;信息损失可以导致意识访问的丢失。选择性注意的研究也显示对信息更丰富的意识刺激的强烈偏爱。

(6) 意识事件的适应性和飞逝的本质。

立即经历感觉输入可以维持到几秒,我们短暂认知的持续存在不到半分钟。相反,庞大的无意识知识可以驻存在长时记忆中。

(7) 内部一致性。

意识以一致约束为特征。一般同时给予两个不一致刺激时,只有一个能变得有意识。当一词多义时,只有一个意义变得有意识。

(8) 有限能力和顺序性。

意识的能力在任何规定的片刻好像限制在仅对一个一致景象,和直接同时观察时脑形成的大量并行处理相反,这样的意识景象流是串行的。

(9) 感觉捆绑。

感觉大脑就其功能作用是分块的,从而不同的脑区对不同的特征(如形状、颜色或者目标运动)作出反应。一个基本的问题是这些就其功能作用分开的脑区怎样协调它们的活动,产生普遍的有意识的综合完形知觉。

(10) 自我特性。

意识经验总是以自我经历为特点,威廉·詹姆士将其称为“观察自我”。自我功能看起来与中央脑区有关,人脑包括脑干、楔前叶(Precuneus)和前额叶(Orbitofrontal)皮层。

(11) 准确可报告性。

意识的大多数使用的行为迹象是准确可报告的。全范围的意识内容因为大范围自愿的反应是可报告的,经常有非常高的准确性。可报告不要求完全明确的词汇,因为主体能自动地对意识事件进行比较、对比、指向和发挥作用。

(12) 主观特性。

意识以事件私有流方式提供给经历主体为特征。这样的隐私没有违反立法。这表明自

我物体综合是有意识认知的关键。

(13) 关注非主流结构。

意识被认为倾向于专注明白清楚的内容,“非主流意识”事件,如亲情感、舌尖经验、直觉等同样重要。

(14) 促进学习。

几乎没有证据表明学习无需意识。相反,意识经验促进学习的证据是压倒一切的,即使隐性(间接的)学习也需要有意识的注意。

(15) 内容的稳定性。

意识内容给人深刻印象是稳定的。例如读者经常使用眼睛运动扫描句子。即使像自身的信念、概念和专题一样的抽象意识内容,可能在几十年内也会非常稳定。

(16) 关注特性。

意识的景象和目标,一般来说是关注外部的来源,虽然它们的形成严重依赖无意识的框架。

(17) 意识知道和决策。

意识对于我们知道周围世界,以及一些我们的内部过程是有用的。意识的表达,包括感觉、概念、判断和信仰,可能特别适于自如的决策。但是,并非全部有意识事件都涉及大范围的无意识设施。这样,意识报告的内容绝不是仅需要被解释的特征。

## 13.3　心理学的意识观

认知科学家主要想用标准的心理学方法对理解意识做出贡献。他们把大脑视为一个不透明的“黑箱”,我们只知道它的各种输入(如感觉输入)所产生的输出(它产生的行为)。他们根据对精神的常识性了解和某些一般性概念建立模型。该模型使用工程和计算术语表达精神[164]。

普林斯顿大学心理系教授约翰逊-莱尔德(Philip Johnson-Laird)是一位杰出的英国认知心理学家。他主要的兴趣是研究语言,特别是字、语句和段落的意义。约翰逊-莱尔德确信,任何一台计算机,特别是高度并行的计算机,必须有一个操作系统用以控制(即使不是彻底的控制)其余部分的工作,他认为,操作系统的工作与位于脑的高级部位的意识之间存在着紧密的联系。

布兰迪斯大学语言学和认知学教授杰肯道夫(Rav Jackendoff)是一位著名的美国认知科学家。他对语言和音乐具有特殊的兴趣。与大多数认知科学家类似,他认为最好把脑视为一个信息加工系统。但与大多数科学家不同的是,他把“意识是怎样产生的”看作心理学的一个最基本的问题。他的意识的中间层次理论认为,意识既不是来自未经加工的知觉单元,也不是来自高层的思想,而是来自介于最低的周边(类似于感觉)和最高的中枢(类似于思想)之间的一种表达层次。他恰当地突出了这个十分新颖的观点。他还认为,意识与短时记忆之间存在紧密的联系。他所说的“意识需要短时记忆的内容来支持”这句话就表达了这样一种观点。但还应补充的是,短时记忆涉及快速过程,而慢变化过程没有直接的现象学效应。谈到注意时,他认为注意的计算效果就是使被注意的材料经历更加深入和细致的加工。他认为这样就可以解释为何注意容量如此有限。

加利福尼亚州伯克利的赖特研究所的巴尔斯教授写了《意识的认知理论》一书,虽然巴尔斯也是一位认知科学家,但与杰肯道夫或约翰逊-莱尔德相比,他更关心人的大脑。他把自己的基本思想称为全局工作空间。他认为,在任一时刻存在于这一工作空间内的信息都是意识的内容。作为中央信息交换的工作空间,它与许多无意识的接收处理器相联系。这些专门的处理器只在自己的领域之内具有高效率。此外,它们还可以通过协作和竞争获得工作空间。巴尔斯以若干种方式改进了这一模型。例如,接收处理器可以通过相互作用减小不确定性,直到它们符合一个唯一有效的解释。广义上讲,他认为意识是极为活跃的,而且注意控制机制可进入意识。我们意识到的是短时记忆的某些项目而非全部。

这三位认知理论家对意识的属性大致达成了三点共识。他们都同意并非大脑的全部活动都直接与意识有关,而且意识是一个主动的过程;他们都认为意识过程有注意和某种形式的短时记忆参与;他们大概也同意,意识中的信息既能够进入到长时情景记忆中,也能进入运动神经系统的高层计划水平,以便控制随意运动。除此之外,他们的想法存在一些分歧。

## 13.4 意识的剧场模型

关于意识问题,最经典的一个假设即所谓"剧场中的亮点"隐喻。在这一个隐喻中,把多个感觉输入综合成一个有意识的经验,比拟为在黑暗的剧场内舞台上有聚光灯打出一个光亮点照到某个地方,然后传播给大量的无意识的观众。在认知科学中,关于意识和选择性注意的假设多数来自于这个基本的隐喻。巴尔斯是"剧场隐喻"的最主要的继承和发扬光大者[41]。

巴尔斯将心理学和脑科学、认知神经科学紧密结合起来,把一个从柏拉图和亚里士多德时代开始就一直被用于理解意识的剧场隐喻改造成意识的剧场模型,并运用大量引人注目的神经影像学的先进研究成果,阐述人类复杂的心灵世界(见图 13.1)[42]。

这一模型的基本观点是:人的意识活动是一个容量有限的舞台,需要一个中央认知工作空间,它与剧场的舞台非常类似。意识作为一种大认识现象的心理状态,基本上有五种活动类型。

(1) 工作记忆就像剧场的舞台,主要包括"内心语言"和"视觉想像"这两种成分。

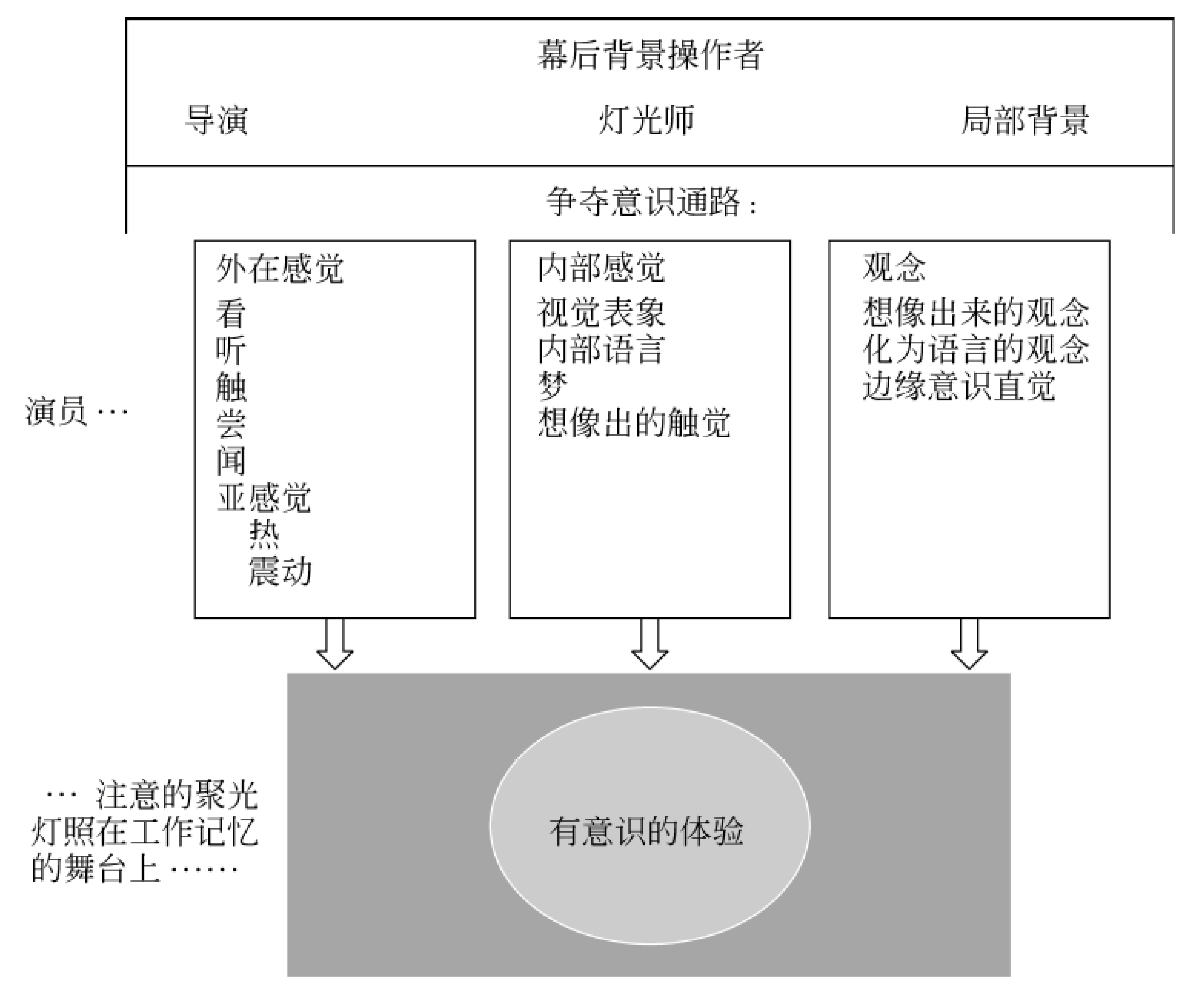
(2) 意识体验的内容好比前台演员,在不同的意识体验内容之间显示出竞争和合作的关系。

(3) 注意如同聚光灯,它照在工作记忆这个舞台上的演员身上时,意识的内容便出现了。

(4) 幕后的背景操作由布景后面的背景操作员系统来执行,其中"自我"类似幕后背景操作的导演,许多普遍存在的无意识活动也构成了类似舞台的背景效应,背景操作员则是大脑皮层上的执行、控制系统。

(5) 无意识自动活动程序和知识资源组成了剧场中的"观众"系统。

按照他的观点,尽管人的意识能力有限,但人的优势却在于可以接触大量的信息资料,并具有某种潜在的计算能力。这些能力包括多感官输入、记忆、先天与后天习得技能等。巴



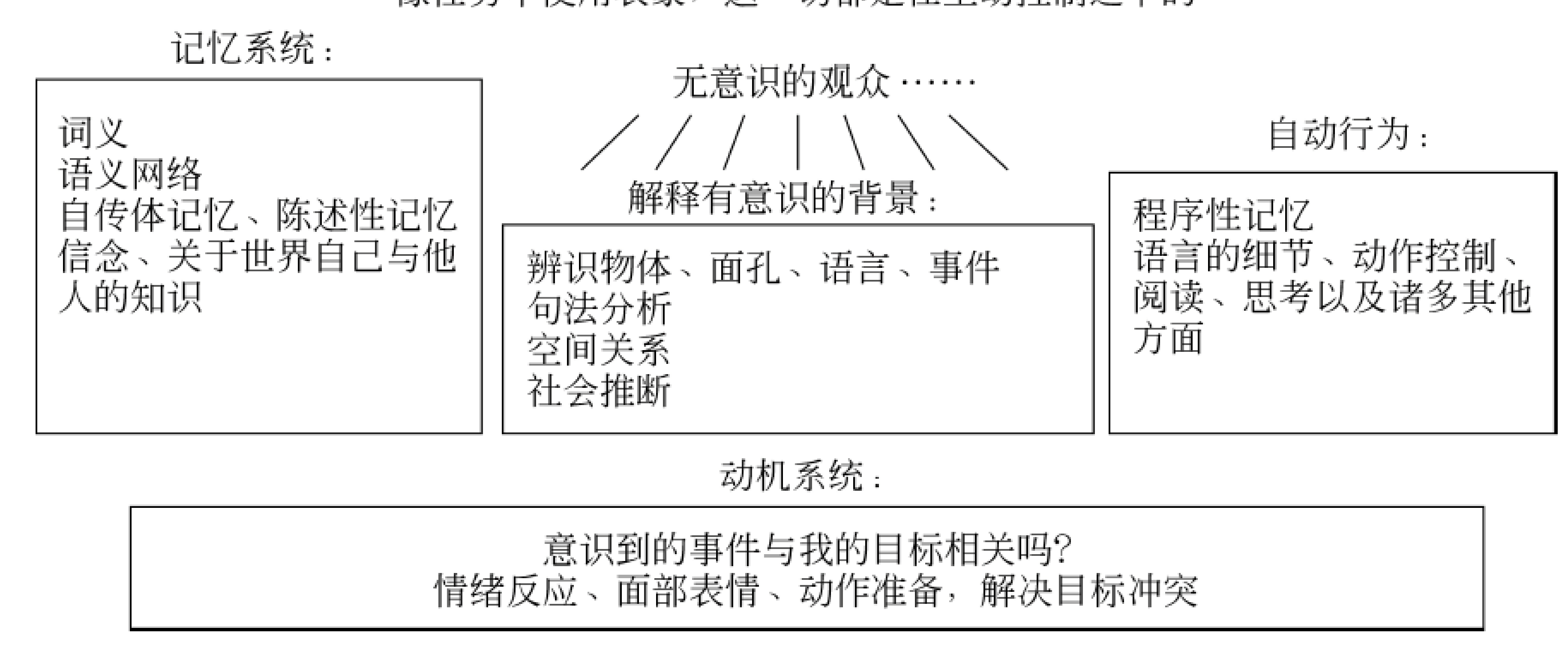




图 13.1 意识的剧场模型

尔斯同时还提出，意识的脑工作是广泛分布式的，就像同时有许多角色在演出的剧场，共有四种脑结构空间维度、四类脑功能模块系统来支撑，它们同时投射在时间轴上，形成一种超立体的空间、时间活动维度的一体化的类似剧场舞台式的心智模型。其中脑结构的四个空间维度同时投射在时间轴上：①从脑的深层到皮层的皮层化维度；②从后头部向前头部发展的前侧化维度；③大脑两半球功能的左右侧化发展维度；④脑背侧和腹侧发展维度，从而组成一种超立体的空间时间维度。

脑功能系统由四类模块组成：①与本能相关的功能模块——具有明确的功能定位；②人类种属特异的本能行为模块——自动化的功能定位；③个体习得的习惯性行为模块——半定位的自动化系统；④高级意识活动——没有明确的定位系统，意识的内容似乎可以整个地传播到遍布大脑的神经网络上，从而形成一个分布式的结构系统。人类意识经验是个统一体，自我是这个统一体的“导演”。

巴尔斯还在“意识剧场模型”的基础上提出“意识与无意识相互作用模型”,简洁地隐喻了意识与无意识之间相互转化的动态过程,即多种形式的意识活动和有意识与无意识活动的相互转化,形成一种复杂的脑内整体工作信息处理、意识内容和丰富多彩的主观自身感受经验约束。根据巴尔斯的观点,在无意识过程建构基础的背后隐藏着一个专门特殊的处理器,功能是统一的或者模块的。特别需要强调的是,无意识处理器十分有效而且快捷,它们很少有错误,同时,这样的处理器可能在操作上与其他系统汇集在一起,专门的处理器是分离和独立的,它们能够对主要的信息进行机动处理。这种专门的处理器的特征十分类似于认知神经心理学上所讲的“模块”。

意识的形成是否由特定的脑过程引起?是否可以用复杂系统来为脑过程的意识形成建立模型?这些是意识研究关心的问题。对于意识活动的神经机制的探索发现,意识的清醒状态是心理活动得以进行的基本条件,而意识的清醒程度明显与脑干的网状结构、丘脑等边缘系统的神经通路存在密切联系。一般来讲,脑干网状结构系统的兴奋性则与注意的强度有关,感官输入的大量信息在经过网状结构系统时需要进行初级的分析整合,许多无关或次要的信息被有选择性地过滤掉,只有引起注意的有关信息才能到达网状结构系统。因此,有学者提出,意识活动主要体现在以网状结构为神经基础的注意机制之上,只有注意到的刺激才能引起我们的意识,而很多非注意的刺激没能达到意识水平就不会被意识到,变成意识的活动依赖一种确定的精神机能——注意的介入。当然,意识与无意识有着不同的生理基础和运行机制。大量的无意识活动是并行处理的过程,而意识活动是串行处理的过程。不同的意识状态可以在非常短的时间内进行快速的转换,意识的开启就是指从无意识状态向意识状态的转化过程[43]。

这一模型比较准确地阐述了意识、无意识、注意、工作记忆和自我意识等的相互联系与区别,也得到了许多神经生物学证据的支持,在学术界的影响越来越大。著名学者西蒙曾说,巴尔斯“为我们提供了关于意识的令人兴奋的解释,将这个问题从哲学的桎梏中解脱出来,将它稳固地置于实验研究的领地之中”[43]。也有的学者认为,巴尔斯的意识剧场模型为当前的意识研究提供了一种核心假设。他比较了无意识与意识心理过程之间的差异,核心思想是讲存在分离的意识与非意识两种有区别的过程。在这样的分析基础上,巴尔斯提出意识与无意识事件是可以被认识的,在神经系统中有着各种不同的建构过程。克里克等人的研究认为,视觉意识产生于大脑枕叶上的皮层投射区,其可能提供了某种“剧场舞台”的探照灯,它会由于注意的激活而照亮起来,从而显示出连贯的意识信息[166]。“观众”是指无意识的脑区,像部分皮层、海马、基底核、杏仁核以及运动执行系统和解释系统。意识的“剧场假设”隐含着在舞台上同时有许多角色在演出,正像人脑同时接受内外感受器的多种刺激,但是只有少量角色接收聚光灯的照射,这中间有个选择问题,而且聚光灯不是停留在一个地方、一个角色身上,而是随着时间流动。

意识的剧场隐喻也受到一些学者的反对。如丹尼特(D C Dennett)认为,这个假设一定要有个“舞台”才能有“意识”演出,那么就是说,大脑中有一个专门地方作为意识的舞台。这种假设很容易落入 17 世纪笛卡儿关于精神的灵魂之源“松果体”假说的案白。反对者认为,大脑中没有一个专门的地方集中所有的输入刺激。

## 13.5 意识的还原论理论

诺贝尔奖获得者,DNA 双螺旋结构的提出者克里克是还原论意识理论的典型代表之一。他认为意识问题是整个神经系统高级功能中的关键问题,所以他于 1994 年出版了一本高级科普书,名为 *The Astonishing Hypothesis*(惊人的假设),副标题为"用科学方法探索灵魂"[164]。他大胆地提出了一个基于"还原论"的"惊人的假说"。他认为"人的精神活动完全由神经细胞、胶质细胞的行为和构成及影响它们的原子、离子和分子的性质所决定"。他坚信,意识这个心理学的难题,可以用神经科学的方法来解决。他认为意识问题与短时记忆和注意的转移有关,他还认为意识问题虽然牵涉到人的许多感觉,但他想从视觉意识着手,因为人是视觉性动物,视觉注意容易进行心理物理实验,而且神经科学在视觉系统研究方面积累了许多资料。20 世纪 80 年代末 90 年代初在视觉生理研究方面有一个重大的发现:从不同的神经元的发放中记录到同步振荡现象,这种大约 40Hz 的同步振荡现象被认为是联系不同图像特征之间的神经信号。克里克等提出视觉注意的 40Hz 振荡的模型。并推测神经元的 40Hz 同步振荡可能是视觉中不同特征进行"捆绑"的一种形式。至于"自由意志,克里克认为它与意识有关;涉及行为和计划的执行。克里克分析了一些"意志"丧失者的情况,认为大脑中负责"自由意志"的部位在于前扣带回,靠近 Brodmanm 区(24 区)。

克里克和科赫(C Koch)认为研究意识的最困难问题是感受性问题,即如何感受到红颜色、痛苦的感觉等。这是由意识的主观性和不可表达性决定的,因而,他们转向研究意识的神经相关物(NCC),即了解意识的某些方面神经活动的一般性质。克里克和科赫列举了意识研究中神经相关物的十条框架[167]。

(1) 无意识的侏儒(Homunculus)。

首先考虑脑整体的工作方式,大脑的前部注视着感觉系统,感觉系统的主要工作是在脑的后部进行的。而人并不直接知道他们的想法,而只知道意象中的感觉表象。这时,前脑的神经活动是无意识的。脑中有一个"侏儒的假设",现在已不再时髦,但是,离开这个假设,人如何想像他们自己呢?

(2) 刻板(Zombie)方式和意识。

对于感觉刺激,许多反应是快速的、瞬态的、刻板的和无意的,而意识处理的东西更慢、更广,且需要更多时间决定合适的想法和更好的反应。进化上发展出这两种策略以相互补充,视觉系统的背侧通道(大细胞系统)执行刻板的快速反应;腹侧系统(小细胞系统)执行的是有意识的识别任务。

(3) 神经元联盟。

此处联盟是 Hebb 集群加上它们之间的竞争。联盟中的神经元并非固定不变,而是动态的。竞争中获得优势的联盟会保持一段时间占据统治地位,这就是我们意识到什么东西的时候。这个过程犹如国家的选举,选举中获胜的政党会执政一段时间,并影响下一阶段的政局。"注意"机制相当于舆论界和选情预测者的作用,试图左右选举形势。皮质第Ⅴ层上的大锥体细胞好像是选票。但是,每次选举之间的时间间隔并不是有规律的。当然这仅仅是比喻。

联盟的大小和特性方面是有变动的。清醒时的意识联盟与做梦时不一样,闭眼想像时

与睁眼观看时也不一样。脑前部分联盟可能反映“快感”“统治感”等自由意志方面的意识,而脑后部的联盟可能以不同方式产生,前后脑的联盟可能不止一个,会相互影响和作用。

(4) 显性表象。

视场中某一部分的显性表象意味着存在一小组神经元,它们对应着这一部分的特性,可以像检测器那样做出反应,而无须复杂的加工。在一些病例中,某些显性神经元的缺失造成某种功能的丧失,如颜色失认症、面孔失认症、运动失认症。这些患者的其他视觉功能仍保持正常。

在猴子实验中,运动皮质(MT/VS 区)一小部分受损,造成运动感知的丧失。损伤部位较少,几天内仍可恢复,若大范围的损伤则造成永久性丧失。必须注意,显性表象是意识的神经相关物的必要条件而非充分条件。

(5) 高层次优先。

一个新的视觉输入来到后,神经活动首先快速地无意识地上行到视觉系统的高层,可能是前脑,然后信号反馈到低层次,所以,达到意识的第一阶段在高层次,再把意识信号发送到额叶皮质,随后在较低层次上引起相应活动,当然这是过于简单的描述。整个系统中还有许多横向联系。

(6) 驱动性和调制性联系。

了解神经联络的本质很是重要,不能认为所有兴奋性联系都是同一类型。可以把皮质神经元的联系粗略地分为两大类:一类是驱动性的;另一类是调制性的。对皮质锥体细胞而言,驱动性联系多半来自基底树突,而调制性输入来自丛状树突,它们包括反向投射,弥散状投射,特别是丘脑的层间核。从侧膝体到 V1 区的联系是驱动性的。从背脑到前脑的联系是驱动性的。而逆向联系多半是调制性的。皮质第五层上的细胞(它投射到丘脑)是驱动性的,而第六层则是调制性的。

(7) 快照。

神经元可能以某种方式超过意识的阈值,或者保持高发放率或某种类型的同步振荡,或者某种簇发放。这些神经元可能是锥体细胞,它投射到前脑。如何维持高于阈值的神经活动呢?这涉及神经元的内部动力学,诸如 $Ca^2$ 等化学物质的积聚,或者皮质系统中再入线路的作用。也可能正反馈环的作用使得神经元的活性不断增加,达到阈值,并维持高活性一段时间。关于阈值问题也可能出现某种复杂性,它可能依赖于达到阈值的速率,或者输入维持多长时间。

视觉觉知过程由一系列静态的快照组成,也就是感知出现在离散的时间内。视皮质上有关神经元的恒定发放率,代表有某种运动发生,运动是发生在一个快照与另一个快照之间,每个快照停留的时间并不固定。对于形状和颜色的快照时间可能碰巧一样,它们的停留时间与 α 节律或 δ 节律有关。快照的停留时间依赖开启信号、关闭信号、竞争和适应等因素。

(8) 注意和绑定。

把注意分成两类是有用的:一类是快速的、显著性驱动的和自下而上的;另一类是缓慢的、自主控制的和自上而下的。注意的作用为了左右那些正在竞争的活跃的联盟。自下而上的注意从皮质第五层的神经元出发,投射到丘脑和上丘。自上而下的注意从前脑出发,分散性地反投射到皮质Ⅰ、Ⅱ和Ⅲ层上神经元顶树突,可能途径丘脑的层间核。普遍认为丘

脑是注意的器官。丘脑的网状核的功能在于从一个广宽的对象中作出选择。注意的作用是在一群竞争的联盟中作出倾向性作用,从而感受到某个对象和事件,而不被注意的对象却瞬间消逝了。

什么是绑定?所谓绑定是把对象或事件的不同方面,如形状、颜色和运动等联系起来。绑定可能有几种类型。如果它是后天造成的或者经验学得的,它可能具体化在一个或几个节点上,而不需要特殊的绑定机制。如果需要的绑定是新的,那么那些分散的基本节点的活动需要联合起来一起活动。

(9) 发放风格。

同步振荡可以在不影响平均发放率情况下增加一个神经元的效率。同步发放的意义和程度仍有争议。计算研究表明其效果取决于输入的相关程度。我们不再把同步振荡(如40Hz)作为神经相关物的足够条件。同步发放的目的可能是在于支持竞争中的一个新生联盟。如果视刺激非常简单,如空场上的一个条形物,此时没有有意义的竞争,同步发放可能不出现。同样,一个成功的联盟达到意识状态,这种发放也可能不必要了。正如你获得一个永久职位后,你可能放松一阵子。在一个基本节点上,一个先期到达的脉冲可能获得的好处大于随后的脉冲。换言之,脉冲的准确时间可能影响到竞争的结果。

(10) 边缘效应和意义。

考虑一小堆群神经元,它们对面孔的某些方面有反应。实验者知道这一小群细胞的视觉特性,但是大脑怎么知道这些发放代表的是什么呢?这就是“意义”问题。神经相关物只是直接关系到所有锥体细胞的一部分,但是它会影响到许多其他神经元,这就是边缘效应。边缘效应由两部分组成,一是突触效应,二是发放率。边缘效应并不是每个基本节点效应的总和,而是作为神经相关物整体的结果。边缘效应包括神经相关物神经元过去的联合,神经相关物期望的结果,与神经相关物神经元有关的运动等。按定义,边缘效应本身不能被意识到,显然它的部分可能变为神经相关物的一部分。边缘效应的神经元的某些成员可能反馈投射到神经相关物的部分成员,支持神经相关物的活动。边缘神经元可能是无意识的启动的部位。

克里克和科赫的意识框架把神经相关物的想法从哲学、心理和神经的角度编织在一起,其关键性的想法是竞争性联盟。猜测一个节点的最小数量的神经元群可能是皮质功能柱。这种大胆的假设无疑给意识的研究指出了一条道路,那就是通过研究神经网络、细胞、分子等各层次的物质基础,最终将找到意识问题的答案。但是这个假设面临着一个核心问题——到底谁有“意识”?如果是神经细胞,那么“我”又是谁?

## 13.6 神经元群组选择理论

诺贝尔奖获得者埃德尔曼依据脑的生理病理和解剖学上的事实,强调意识的整合性和分化性[204]。他认为丘脑—皮质系统在意识的产生方面起关键作用。这里丘脑特指丘脑层间核,网状核和前脑的底部,统称为“网状激活系统”,这部位的神经元弥散性地投射到丘脑和皮质,它的功能是激发丘脑-皮质系统,使整个皮质处于清醒状态。近年来的一些无损伤实验表明,皮质的多个脑区同时激发,而不是单一脑区的单独兴奋。

2003 年,埃德尔曼在美国科学院系列(PNAS)上发表一篇论文[203],一开始就主张掘弃二元论。他分析了意识的特性后,指出意识研究必须考虑如下几方面。

(1) 意识状态的可变性,分化性与联合统一后出现个体性之间的反差,其统一性又需要把来自各感觉通道的信息绑定在一起。

(2) 意向性,表明意识是一般的,同时,意识又受注意调制,并与记忆和意象有广泛的联系。

(3) 主观感觉和感受性。神经科学表明,意识不是单个脑区或某些类型神经元的性质,而是广泛分布的神经元群体(Group)中动态相互作用的结果。对意识活动起主要作用的系统是丘脑-皮质系统。意识经验的整合动态性认为丘脑-皮质系统的行为像一种功能性簇(Cluster),其相互作用主要发生在其本身,当然,与其他系统也有一些相互作用,例如,与基底核的相互作用。在这些神经结构中活动的阈值受到上行价值系统的支配,如中脑的网状系统与丘脑层间核的相互作用,去甲肾上腺素能、五烃色胺能、胆碱能和多巴胺能核团。丘脑掌控了意识状态的水平,来自层间核的输入改变皮质活动的阈值。此外,在睡眠时,脑干对丘脑的作用影响意识状态起重要作用。

埃德尔曼认为脑是一个选择性系统,后天产生大量可变的线路,在经验中选择出某个特殊的线路。在这个选择系统中,结构不同的线路可能进行相同的功能或产生相同的输出。这就是神经元群组选择理论(Theory of Neuronal Group Selection,TNGS)。在这个理论中有一重要概念,即再入(Re-entry),它是一个过程,也是意识涌现的中心环节。这是埃德尔曼一贯主张的观点,再入是脑皮质内区域之间众多平行互逆纤维中进行的循环信号。再入是平行进行的选择性过程,它不同于反馈,后者是指令性的,涉及误差函数,而且是信号通道中序列式传递。竞争性神经元群组之间相互作用加上再入,在广泛分布的脑区中的同步活动,都会由于再入而决定选择的取向。这也可能为绑定问题提出一个解决方案,即在缺少操作程序和上级协调者的情况下,如何把不同脑区的活动相关起来。把功能上分离的脑区活动联系起来是感觉分类的一个中心问题。

按神经元群组选择理论,脑中选择性事件受到上行的弥散的价值系统的约束。价值系统用调制或改变突触阈值的办法影响到选择过程。价值系统包括蓝斑(Locus Coeruleus)、缝核(the Raphe Nucleus)、胆碱能、多巴胺能、组胺能核。边缘系统和脑干价值系统会作用到突触强度的改变。这一系统极大地影响前脑、顶叶和颞叶皮层的活动,也是意识涌现的关键。

埃德尔曼提出的神经元群组选择理论(或神经达尔文主义)是他的意识理论框架的中心,主要体现在以下两点:①从本质上来说,一个选择性神经系统有十分巨大的多样性,这一点是脑意识状态复杂性所必需的;②再入在此起关键作用,它把分散的多个脑区的活动联系起来,然后在感觉分类时动态地改变。因此,多样性和再入两者是意识经验的基本性质。

埃德尔曼把意识分为两类:一类是初级意识;另一类是高级意识。初级意识只考虑眼下的事件,高级意识只是在进化的后期才出现,在人类达到最高级阶段,可以使用语言交流,并可对行为做出计划。但神经活动在这两类意识中应当是类似的。埃德尔曼认为爬行类进化到哺乳类,大量新的互逆性联系发展出来,使得丰富的再入活动在前后脑之间发生,而后脑主要对感觉分类负责,前脑对价值系统负责。这种再入活动为感觉综合提供神经基础,也

为眼前的复杂场景与过去经历的事件的记忆进行联系。在进化的最后期，再入通路把语义和行为联系起来，并形成概念。从而出现高级意识。

在此基础上，埃德尔曼引入"再入性动态核心"的概念。在一个复杂系统中，由许多小区域组成，它们之间半独立地活动，又通过相互作用形成较大的集群以产生整合性功能。丘脑-皮质系统就是这种复杂系统。再入性动态核心是一种过程，在 500ms 或少于这个时间内形成一种功能簇堆，然后向其他再入性动态核心转移，再入性动态核心就是功能簇。这发生在复杂系统中，以产生多样化的统一状态，这一点与克里克的"竞争性联盟"有许多共同之处。

## 13.7 意识的量子理论

量子论揭示了微观物质世界的基本规律，是所有物理过程、生物过程和生理过程的微观基础。量子系统超越了粒子与波或相互作用与物质的分别，以不可分割的并行分布式处理综合起作用的。非局域性和远距相关性是量子特性，量子整体可能与意识密切相关。

量子波函数坍缩是一种变迁，指量子波函数从众多量子本征态线性组合的描述态向一个本征纯态的变迁，简单地说就是众多量子图式的迭加波变换成单一的量子图式。波函数坍缩意味着一种从亚意识记忆到显式记忆的意识表象的选择性投射。有两种可能的记忆和回忆的理论，上面提到的量子理论，或经典（神经）理论，记忆可能是突触连接系统的一种并行分布图式，但也可能是更精细的结构，如由埃弗里特（H Everett）提出的多世界解释量子理论的并行世界及玻姆的隐次序等[215]。

澳大利亚国立大学脑意识研究中心主任、哲学家查尔默斯（David Chalmers）提出了多种量子力学方式来解释意识[131]。他认为，坍塌的动力机制为相互作用论者的解释提供了开放余地。查尔默斯认为问题在于我们如何解释。我们想知道的不仅仅是关联，我们想要解释大脑过程如何产生意识、为什么产生意识，这才是神秘之处。最有可能的解释是，在意识状态不可能被叠加的条件下，意识状态和系统的整体量子状态有关。大脑作为意识的物理系统，在非叠加的量子状态中，该系统的物理状态和精神现象相互关联。

美国数学家和物理学家彭罗斯（Roger Penrose）从歌德尔定理发展了自己的理论，认为人脑有超出公理和正式系统的能力。他在第一部有关意识的书《皇帝新脑》中提出[552]，大脑有某种不依赖于计算法则的额外功能，这是一种非计算过程，不受计算法则驱动；而算法却是大部分物理学的基本属性，计算机必须受计算法则的驱动。对于非计算过程，量子波在某个位置的坍塌，决定了位置的随机选择。波函数塌缩的随机性，不受算法的限制。人脑与电脑的根本差别，可能是量子力学不确定性和复杂非线形系统的混沌作用共同造成的。人脑包含了非确定性的自然形成的神经网络系统，具有电脑不具备的"直觉"，正是这种系统的"模糊"处理能力和效率极高的表现。而传统的图灵机则是确定性的串行处理系统，虽然也可以模拟这样的"模糊"处理，但是效率太低下了。而正在研究中的量子计算机和计算机神经网络系统才真正有希望解决这样的问题，达到人脑的能力。

彭罗斯又提出了一种波函数塌缩理论，适用于不与环境相互作用的量子系统，却可能自行塌缩。他认为，每个量子叠加有自身的时空曲率，当它们距离超过普朗克长度（$10^{-35}$m）时就会塌缩，称为客观还原（Objective Reduction）。彭罗斯认为，客观还原所代表的既不是

随机,也不是大部分物理所依赖的算法过程,而是非计算的,受时空几何基本层面的影响,在此之上产生了计算和意识。

1989年彭罗斯在撰写第一部关于意识的书《皇帝新脑》时,还缺乏对量子过程在大脑中如何作用的详细描述。从事癌症研究和麻醉学的哈梅罗夫(S R Hameroff)读了彭罗斯的书,提出了微管结构以支持大脑量子过程。支持神经元的细胞骨架蛋白主要由一种微管构成,而微管由微管蛋白二聚体亚单元组成,其功能包括传输分子、联系神经突触的神经传导素、控制细胞生长等。每个微管蛋白二聚体都有一些憎水囊,彼此间距约8nm,里面含有离域π电子。微管蛋白还有更小的非极性域,含有π电子富集吲哚环,相隔约2nm。哈梅罗夫认为这些电子之间距离很近,足以形成量子纠缠。

哈梅罗夫进一步提出,这些电子能形成一种玻色-爱因斯坦凝聚态,而且一个神经元中的凝聚态能通过神经元之间的间隙接点扩展到其他多个神经元,由此在扩展脑区形成宏观尺度的量子特征。当这种扩展的凝聚波函数坍塌时,就形成了一种非计算性的影响,而这种影响与深植于时空几何中的数学理解和最终意识体验有关。而这种凝聚态的活动性造成了大脑中的伽马波同步,传统神经科学认为这种同步与意识和间隙接点的功能有关。

彭罗斯和哈梅罗夫合作,在20世纪90年代早期共同建立了广受争议的"和谐客观还原模型(Orch-OR模型)"。按照Orch-OR规定的量子叠加态进行运算之后,哈梅罗夫的团队宣布新的量子退相干所需的时间尺度要比泰格马克(Max Tegmark)的结果大7个级数。但这个结果依然比所需的时间少了25ms——如果想要使量子过程如同Orch-OR所描述的那样,能够和40Hz的伽玛同步产生关联。为了弥补这一环节,哈梅罗夫等人做了一系列假设和提议。首先他们假设微管内部可以在液态和凝胶态之间互相转换。在凝胶状态下,他们进一步假设水的电偶极子会沿着微管外围的微管蛋白同向排列。哈梅罗夫认为这种有序排列的水将会屏蔽微管蛋白中任何量子退相干过程。每个微管蛋白还会从微管中延伸出一条带负电荷的"尾巴",从而可以吸引带正电荷的离子。这可以进一步屏蔽量子退相干的过程。除此之外,还有推测认为微管可在生物能的驱使下进入相干态。

佩罗斯(M Perus)将神经计算与量子意识相结合的设想[561]。在神经网络理论中神经元系统的状态是由一个向量描述的,正好反映的是神经元系统的随时间变化的活动性分布。特定的神经元图式代表一定的信息。在量子理论中量子系统的状态则可以用随时间变化的波函数描述。这样一来,神经元状态是神经元图式的一种叠加,就可以变为是量子本征波函数的一种迭加了,并且迭加的量子本征波函数通常具有正交性和正则性。在本征态的线性组合中,每一种本征态有一个对应的系数,描述在系统的实际状态中一种特定意义表达的可能性程度。神经元信号的时空整合可以用薛定锷方程的Feynman形式来描述。神经系统从潜意识到意识转变对应到"波函数坍缩",是从隐序到显序转变的结果[89]。神经系统模型是以显式方式来给出神经系统的空间信息编码的,而对于时间信息编码则要来得间接。不过通过傅里叶变换,我们同样很容易建立起具有显式时间结构信息的描述方程。如果神经激活图式代表意识对象的描述,那么傅里叶变换后的神经激活频谱,代表的神经元激活振荡频率分布就与意识本身相关联了。这就是意识活动互补性的两个方面,共同给出意识过程整体性时空编码。

## 13.8　综合信息理论

托诺尼(G Tononi)与埃德尔曼等发表了一系列论文,阐明意识的综合信息理论[758]。文献[56,759]提出意识量是由复杂元素生成的综合信息量,并由它生成的信息关系规定的体验质量。托诺尼提出综合信息的两个测度[13]。

### 1. 测度$\Phi_1$

神经系统静态性质的度量。如果托诺尼是正确的,它会测量在一个系统中类似的意识潜力。它不能是系统当前的意识水平,因为它是一个固定的神经结构的固定值,不管系统当前的发放率(例如,响应于输入或内部动态变化)。托诺尼的第一项测度的工作原理是考虑所有的各种双分区的神经系统(分裂成两部分):综合信息的能力被称为 $\Phi$,并且由双分区子集可以交换的最小有效信息给定。托诺尼的方法需要检查所考虑的系统每个子集。每个双分区分为两个不重叠的部分。假设子集 S,可二分为 A 和 B,托诺尼定义了一个测度,称为有效信息(EI)。有效信息使用信息论中的标准度量互信息(MI)。这不是标准的互信息测度,而是考虑 A 和 B 之间的连通性的信息增益互信息测度。托诺尼的 EI 是一个衡量累积的信息增益测度,当 A 的输出在所有可能的值随机变化时,考虑对 B 的效果。其目的是将因果关系的一些因素结合起来。互信息 MI 可以用下面公式描述:

$$\mathrm{MI}(A:B)=H(A)+H(B)-H(AB)$$

其中,H(…)是熵,反映不确定性的测度。如果 A 和 B 之间没有交互,则互信息为零,否则它是正值。

### 2. 测度$\Phi_2$

托诺尼和合作者提出 $\Phi$ 的修订测度,那就是 $\Phi_2$。该修订测度 $\Phi_2$ 比前面的测度优越,因为它可以处理随时间变化的系统,提供一个瞬时到瞬时变化的 $\Phi_2$ 测度,对应于衡量瞬时到瞬时的意识水平。

$\Phi_2$ 也被定义为有效信息,但是有效信息现在的定义与 $\Phi_1$ 版本的完全不同。在这种情况下,有效信息是通过已知的因果结构中,系统在离散的时间步长下演变定义。考虑系统在时间 $t_1$ 时的状态 $x_1$。给定该系统的体系结构,只有某些状态可能导致 $x_1$。托诺尼称这种状态的集合(其相关概率)为后验项。托诺尼还需要一个系统可能状态(和它们的概率)的测度,在这种情况下,我们不知道时间 $t_1$ 时状态的情况,托诺尼称之为先验项。在我们不知道所有关于它的因果架构的情况下,必须把每一个神经元的所有可能的激活值看成同样可能,计算先验项。先验和后验项将有各自相应的熵值。例如,如果先验项包括四个同样可能的状态,和后验项有两个同样可能的状态,那么熵值将分别为两比特和一比特。这意味着,在时间 $t_1$ 发现系统的状态为 $x_1$,获得了较早一个时间步的系统状态的信息。

托诺尼认为,这是系统变成状态 $x_1$ 时生成多少信息的测度。在定义该系统有多少信息生成的测度时,托诺尼再次要求如何"整合"这个信息测度。因此,他观察了可以任意地分解系统的可能性。对于每个部分(单独考虑)给定的当前状态只能来自某些可能的父状态。因

此,我们可以问,有没有可能分解成几部分,使系统整体的信息大于单独的部分的信息? 如果有可能,那么我们已经找到一种方法来将系统分解成完全独立的部分。

在系统不能分解成完全独立的部分的情况下,我们可以寻找整体相对于部分最低的附加信息的分解。托诺尼称这是最小信息划分。最小信息划分的有效信息(由整个系统给定的附加信息,而不是部分)是该系统的 $\Phi_2$ 值。

最后,我们通过对所有的子系统和所有的分区进行穷举搜索来定义复杂性。复杂性是系统具有给定的 $\Phi_2$ 的值,这不包含在任何具有较高 $\Phi$ 的大系统内。类似地,整个系统的主复杂性用最高的 $\Phi_2$ 复杂性表示,系统 $\Phi_2$(或意识)的真正测度是主复杂性的 $\Phi_2$。

在研究 $\Phi_2$ 时,我们注意到,很多 $\Phi_1$ 的问题仍然适用。EI 和 $\Phi_2$ 本身在这方面是紧密联系在一起的,特别是检查特定的系统时。虽然 $\Phi_1$,$\Phi_2$ 和 EI 是通用的概念,目前的数学没有这样广泛适用的标准信息论测度。对于信息综合理论的不足之处的进一步讨论,可以参考文献[64]。

托诺尼认识到,信息综合理论被用来研究系统维持状态的能力,可以说是"智能"。他在文献[758]中描述这种状态质量的方法,并与巴勒杜兹(D Balduzzi)仔细推敲了感受性(Qualia)[56]。感受性原来主要用于哲学家,以便说明内部体验的质量,如玫瑰的红色。

托诺尼宣布已经找到感受性的信息机制,勇敢地面对周围的争议。托诺尼以几何的方式引入形状,体现由系统相互作用产生的一整套信息关系作为感受性的概念。文献[56]探讨了感受性涉及底层系统的特征和体验的基本特征,提供关于感受性几何神经生理学和现象学几何的初始数学词典。感受性空间(Q)是每个具有复杂性可能状态(活动模式)的轴线空间。在 Q 内,每个子机制规定一个点对应系统状态。在 Q 内项目之间的箭头定义信息的关系。总之,这些箭头规定感受性的形状,反映意识体验的质量具有完全和明确的特点。形状的高度 W 是与体验相关的意识量。

## 13.9 显意识思维与潜意识思维

了解了意识的本质就不难在此基础上理解什么是显意识思维和潜意识思维。狭义而言,意识是人脑对时间逻辑思维和空间结构思维(包括形象思维和直觉思维)的觉察、调节或控制。要判断是否有意识产生,至少要能"觉察"出思维过程(觉察出来后,才有可能在此基础上加以调节或控制)。任何思维过程(不管是何种形式的思维过程)都离不开四个要素:思维加工对象(即思维的材料)、思维加工的手段或方法(如分析、综合、抽象、概括、判断、推理、想像等)、思维加工缓存区(也叫"工作记忆",用于暂存思维加工对象及加工结果)和思维加工机制。换言之,只要缺少其中任一要素,思维过程都将因难以进行而不能被觉察出来。因此,如果只是要判别是有意识或无意识即仅局限于觉察思维过程,问题就简单得多,因为要觉察出某种思维过程,并不需要去仔细分析四个要素中的每一个要素,只要抓住其中一个最简单又与觉察直接相关的要素就行了,这个要素就是"工作记忆"(思维加工的缓存区)。这是因为,工作记忆与长时记忆不同,工作记忆中的内容并不长期保留,它只起缓冲存储器作用:在思维加工过程中暂存加工对象和加工结果,加工完成后,其中内容就会很快消失。因此,我们不必考虑思维加工的方式、机制,也不必管思维加工的对象,只需考虑工作记忆中是否有内容(有或无,至于有什么样的内容则不必管),以及工作记忆持续时间的长短就可以

觉察出是否有思维过程发生。

经过多年，特别是近十年脑科学的研究进展，已经发现思维过程涉及两类不同的工作记忆：一类用于存储言语材料（概念），采用言语类编码；另一类用于储存视觉或空间材料（表象），采用图形编码。进一步的研究表明，不仅概念和表象有各自不同的工作记忆，而且表象本身也有两种不同的工作记忆。这是因为，事物的表象有两种：一种是表征事物的基本属性，用于对事物进行识别的表象，一般称为"属性表象"或"客体表象"；另一种是用于反映事物空间结构关系（与视觉定位有关）的表象，一般称为"空间表象"或"关系表象"。空间表象不包含客体内容的信息，只包含确定客体空间位置或空间结构关系所需的特征信息。这样，我们就有三种不同的工作记忆。

(1) 存储言语材料的工作记忆（简称言语工作记忆）：适用于时间逻辑思维。

(2) 存储客体表象（属性表象）的工作记忆（简称客体工作记忆）：适用于以客体表象（属性表象）作为加工对象的空间结构思维，即通常所说的形象思维。

(3) 存储空间表象（关系表象）的工作记忆（简称空间工作记忆）：适用于以空间表象（关系表象）作为加工对象的空间结构思维，即通常所说的直觉思维。

当代脑神经科学的研究成果已经证明，这三种工作记忆以及它们各自对应的思维加工机制，均可在大脑皮层中找到各自对应的区域（尽管有些工作记忆的定位目前还不很准确）。

根据目前脑科学研究的新进展，布朗大学的布隆斯腾指出，言语功能并不是定位在一个狭小的区域上（按传统观念，言语功能只涉及左脑的布洛卡区和沃尼科（Wernicke）区），而是广泛地分布于左脑外侧裂周围区域上，并向额叶前部和后部延伸，包括布洛卡区、紧邻脸运动皮层的下额叶和左侧中央前回（但不包括额极和枕极）。其中布洛卡区受损将影响言语表达功能，沃尼科区受损将影响言语理解功能。但是和言语理解与表达有关的加工机制并不仅限于这两个区（而在此之前，认为由这两个区完全决定人类言语功能的传统观念，在心理学界统治长达一个多世纪）。用于暂存言语材料的工作记忆一般都认为是在"左前额叶"，但具体是在左前额叶中的哪一部位，目前尚未精确定位。据皮特瑞兹（Petrides）等人的意见，倾向于在左前额叶的布洛德曼 6 区[564]。

与言语工作记忆相比，客体工作记忆与空间工作记忆的定位情况要准确得多。1993 年密歇根大学心理系的乔耐德（John Jonides）等人运用当代研究脑科学的最先进测量技术之一，正电子发射断层扫描术（简称 PET），对客体表象与空间表象的生成过程作了深入研究，得到了关于这两种表象生成机制与工作记忆定位的、富有价值的成果。由于 PET 是通过发射正电子的同位素作为标记物，将其引入脑内某一局部区域参与已知的生化代谢过程，然后用计算机断层扫描技术，将标记物参与代谢过程的代谢率以立体成像形式表达出来，因此具有定位准确、对大脑无损伤、适合于大量被试进行测试的优点。

客体表象的生成与加工机制在大脑左半球的颞下回（集中于布洛德曼 37 区，立体定向坐标为：48，－58，－11）、左半球的顶叶（集中于布洛德曼 40 区，立体定向坐标为：35，－42，34）和右半球的扣带回前部（集中于布洛德曼 32 区，立体定向坐标为：－1，14，43）；客体工作记忆则在左半球的前额叶皮层（集中于布洛德曼 6 区，立体定向坐标为：39，3，29）。空间表象的生成与加工机制在大脑右半球的枕叶（集中于布洛德曼 19 区，立体定向坐标为：－30，－76，31）顶叶后部（集中于布洛德曼 40 区，立体定向坐标为：－42，－40，36）和运动前区（集中于布洛德曼 6 区，立体定向坐标为：－34，－1，45）；空间工作记忆则在右半球的

前额叶皮层(集中于布洛德曼47区,立体定向坐标为:−35,19,−2)。

应当说明的是,在上述各组测试结果中,当某一半球有显著激活时,另一半球的相应位置也有激活,只是这种激活未达到统计学上的显著标准而已。

上述结果表明,与客体识别有关的视觉信息加工机制主要在左半球(只有"扣带回前部"不在左半球),客体工作记忆也在左半球;与空间表象有关的视觉信息加工机制,包括空间工作记忆内在,全都在右半球。

应当指出的是,乔耐德采用的是对大脑无损的PET技术,其测试对象不是脑部受损病人而是自愿参试的一批普通大学生(空间表象作业被试18人,客体表象作业被试12人)。实验记录包括6种扫描,每种扫描进行20次实验;每个被试的PET图像都转换成一种立体定位坐标,转换时要求出被试在某种条件下的平均数,并将被试数据进行标准化,然后用Bonferroni法对结果进行多重比较校正。因而其测试结果是比较科学的、可信的。

时间逻辑思维是建立在语言符号序列的基础上,所以具有顺序性、持续性的特点。也就是说,在运用基于语言符号的概念来进行判断、推理时,只能按照词语符号出现的先后顺序一步一步地操作,持续时间往往较长,尤其是在分析处理事物之间的复杂关系时更为突出。总之,在逻辑思维情况下,其工作记忆的持续时间是较长的,加上是按线性、顺序方式工作,每一步骤都很清楚、明确。因此这种思维过程很易于觉察,而且还可以由思维主体将思维过程的每一步骤用言语表述出来,使内隐的心理操作过程变成外显的言语活动。正是在这个意义上,通常把时间逻辑思维称为"显意识思维"。换言之,显意识思维是指其过程不仅能被觉察还能用言语加以描述的思维。

空间结构思维的特点则不太一样。空间结构思维的材料(加工对象)是表象(而不是言语概念),如上所述,表象又分事物表象(客体表象)和空间表象两种,两种表象的加工特点也不相同。

对于事物表象(包含事物基本属性的信息,用于识别不同的事物)的加工,通常是用分析、综合、抽象、概括、想像(又分再造想像和创造想像)等方法。每一方法的具体操作过程的步骤比用言语说明更明确。但是由于在空间结构思维的场合,是以完整的表象作为思维加工单位(而不是以一个个零碎的词语作为加工单位),因而在这种场合,尽管在对表象进行分析、综合、抽象、概括、想像等过程中(通常把这样的思维加工过程也称作"形象思维"),也是按顺序一步一步操作,但由于加工单元少,步骤较简单,所以工作记忆持续时间往往较短,有时在一瞬之间对事物表象的分析、综合就已完成。这种思维也能划分出心理操作步骤,故也能用言语描述,因而具有显意识思维的特征。也就是说,在以事物表象作为加工对象的空间结构思维场合(即通常所说的形象思维场合),一般属于显意识思维,但是有两种情况应该除外。

(1) 在工作记忆持续时间过短(比如只有不到一秒钟),而思维主体神经中枢又为事先分配注意的情况下,这种很短暂的思维过程可能未被觉察,从而成为潜意识(或称"无意识")思维。

(2) 在"创造想像"的酝酿阶段,一般也有一段潜意识思维过程。由于创造想像和再造想像不同,它没有现成表象可以利用,要"无中生有",重新创造出一个前所未有的新表象来,因此在这样的新表象被初步构想出来之前,在工作记忆中将会有一段空白。这段工作记忆为空白的时间,由于缺乏思维加工对象,一般说来思维过程将不能进行,也无法被觉察。更

不能用言语描述，所以这是真正的潜意识思维过程。但是，如果这段时间内有时间逻辑思维与之配合，则这种思维过程有可能展开，从而使创造想像过程得以完成。时间逻辑思维与创造想像过程是相互配合的。

和对事物表象的加工相比，对于空间表象（用于空间视觉定位）的加工有许多新的特点。除了是用完整表象作为思维加工单位，这一点与事物表象的加工有共性以外，其他方面皆与对事物表象的加工不同：它不能通过分析、综合、抽象、概括、想像等手段来对空间表象进行一步一步的顺序结构关系作出判断；它着重考虑的是事物之间的关系（空间位置关系，或其他结构关系），而不是每个事物的具体属性。总之，这是与时间逻辑思维不同，也与以事物表象为加工对象的空间结构思维（即形象思维）不同的另一种空间结构思维。它不是线形、顺序的慢节奏加工，而是在整体综合、直观透视基础上快速作出的直觉判断。所以一直也把这种思维（即以空间表象作为加工对象的思维）称为“直觉思维”。

直觉思维通常有两种：简单直觉思维与复杂直觉思维。简单直觉思维的加工对象（思维材料）是与空间视觉定位有关的空间位置表象，即上面所述的空间表象；复杂直觉思维的加工对象则是用来描述复杂事物之间结构关系的“关系表象”（关系表象是空间表象的一个子类，其全称应是“空间结构关系表象”，与“空间位置表象”同属空间表象的两个子类，但目前已习惯把空间位置表象简称为“空间表象”，“空间结构关系表象”则简称为“关系表象”）。在空间视觉定位情况下（即简单直觉思维过程中），工作记忆内必有关于客体位置的初始特征值，以便根据这些特征值由思维加工机制确定客体的空间位置；而在判断、处理复杂事物关系的情况下（即复杂直觉思维过程中），由于事物之间隐含的复杂关系是有待发现的，所以在工作记忆中将不会有初始值。这是两种直觉思维的很大不同之处。

此外，还应注意不要把直觉思维理解为仅凭直观感觉、没有什么道理、没有任何依据的思维，更不是主观的臆想，而是建立在坚实的理论基础、丰富的实践经验、深入的调查研究和敏锐的观察力与高度的概括力基础上的快速思维。这是因为，如果不具备理论、经验、调研、观察与概括等诸方面的条件，就决不可能在一瞬之间看出全局性的复杂问题或复杂关系中的内在联系，并抓住其中的主要关键，从而“成竹在胸”迅速作出较准确的判断。当然，直观判断毕竟没有经过严密的逻辑分析与推理，因而有时难免不够全面，甚至可能错误，所以在有较充分时间的条件下，最好还是应该运用时间逻辑思维加以验证，以确保无虞。

由以上分析可见，以空间表象作为加工对象的空间结构思维（即直觉思维），由于其特点是整体综合、直觉判断的快速思维（而非线性、顺序、逐步分析的缓慢思维），其工作记忆必然短暂，因而，这种思维过程是比较难以觉察的；加上其思维过程一般没有明确步骤，也就难以用言语来描述。换言之，这种直觉思维若不专门给予注意是不容易觉察的，因而往往表现为“潜意识思维”。尤其是在复杂直觉思维中，由于事物之间复杂的内在关系一时很难把握住，甚至经过较长时间思索也找不到这种关系。这时将会出现和创造想像过程中的类似现象，有一段时间工作记忆的内容为空白。与创造想像过程的区别只在于：创造想像中要建构的是前所未有的新事物的表象，而在复杂直觉判断中则是要发现他人从未揭示过的事物之间的某种隐蔽的关系。可见复杂的直觉思维是和创造想像一样的真正的潜意识思维，即使神经中枢事先给予充分注意（有预期），也无法觉察出其思维过程，更无法用言语去描述该过程。

通过上述讨论，关于显意识思维和潜意识思维可得到以下认识。

(1) 时间逻辑思维,由于其思维过程易于觉察,且可用言语描述,故是显意识思维。

(2) 空间结构思维与显、潜意识的关系,呈现多种不同的情况。

下面列出空间结构思维的四种情况。

(1) 以事物表象(即客体表象)作为加工对象时(也称"形象思维"),其思维过程一般可以被觉察,也可以用言语描述,所以也属于显意识思维(只有工作记忆特别短暂且未加注意,或是处于"创造想像"阶段这两种情况除外)。

(2) 以空间表象作为加工对象时(也称"简单直觉思维"),其思维过程一般不易觉察,且难以用言语描述,所以本应属于潜意识思维,但是假如事先能予以注意,并无法将空间视觉表象转换为时间轴上的事件序列,则这种思维也可转化为显意识思维。

(3) 以关系表象作为加工对象时(也称"复杂直觉思维"),由于这种思维难以觉察(即使给予充分注意,仍无济于事),也不能用言语描述,故是真正潜意识思维。

(4) 在形象思维的两种例外中,第一种要在一定条件下(即工作记忆短暂且为给以注意)才能成为潜意识思维,而第二种(创造想像)才是无条件的真正的潜意识思维。

根据以上各种不同情况,可归纳出下列结论。

(1) 时间逻辑思维是无条件的显意识思维,形象思维一般来说也属显意识思维。

(2) 复杂直觉思维和创造想像皆属无条件的潜意识思维。

(3) 形象思维在一定条件下可以是潜意识思维,简单直觉思维在一定条件下可以是显意识思维。

由于音乐、绘画和文学创作这类创造性活动主要取决于"创造想像"和"时间逻辑思维"的结合,而自然界和人类社会中各种运动变化规律的发现(即自然科学和社会科学的理论探索)这类创造性活动则主要取决于"复杂直觉思维"和"时间逻辑思维"的结合,而创造想像和复杂直觉思维皆属无条件的潜意识思维,时间逻辑思维则是无条件显意识思维。可见,对于创造性思维来说,真正具有决定意义的是无条件的显意识思维和无条件的潜意识思维。所以,今后当我们谈到显意识和潜意识思维时,若无特别说明,皆是指这两种无条件的情况。

## 13.10 机器意识系统

图13.2给出了心智模型CAM的机器意识系统,它由觉知模块、全局工作空间、注意模块、动机模块、元认知模块、内省学习模块构成。

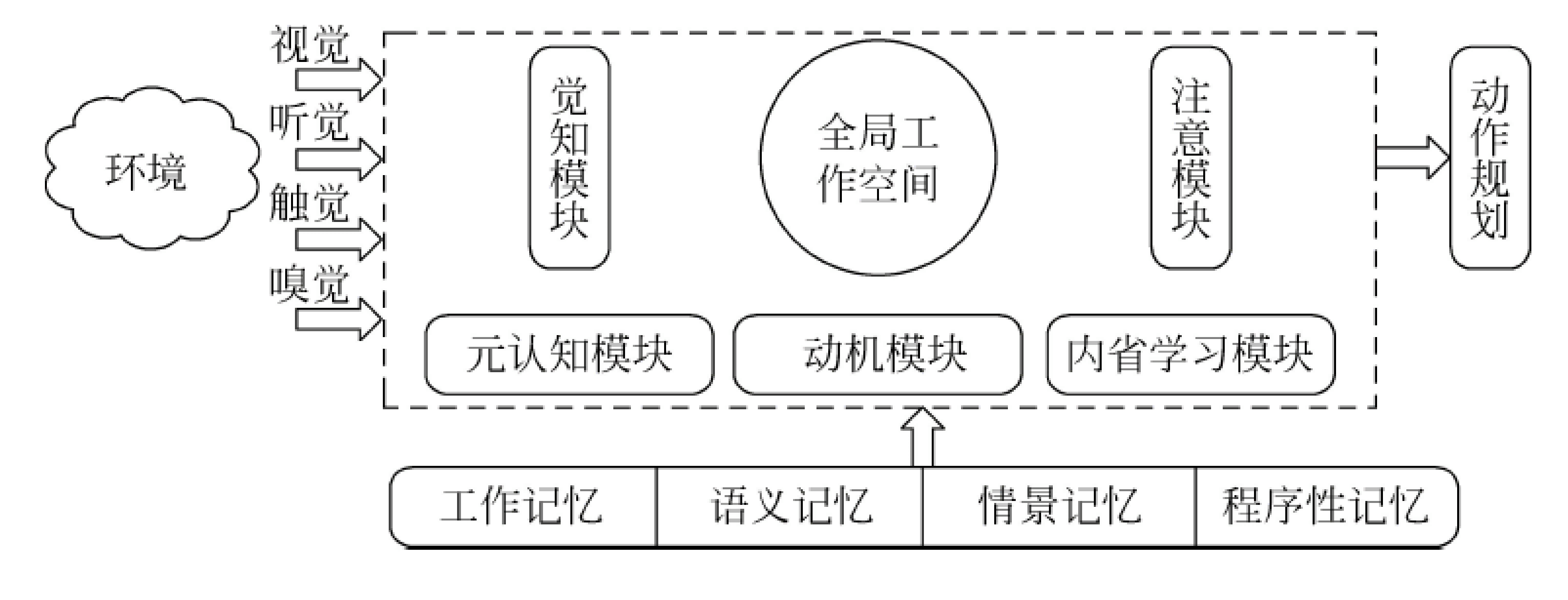



图13.2 CAM的意识系统

觉知模块开始于外界刺激的输入，激活感知系统的初级特征检测器。输出信号被发送到感觉记忆中，在那里更高层次的功能探测器用于更抽象的实体，如对象、类别、行动、事件等的检测。所产生的知觉移动到工作区，在那里产生本地联系的短暂情景记忆和陈述性记忆会被做线索标记。这些本地联系与知觉结合，产生当前情景模型，用以表示智能体对当前正在发生的事情的理解。

全局工作空间模块是处在工作记忆部位，在这个记忆里不同的系统可以执行它们的活动。全局意味着这个记忆中的符号通过众多的处理器被分配、传递开来。当然，每一个处理器都可能产生一些局部的变量并运行。但它对全局性的符号、信息却是相当敏感，可以及时做出感应。当面对全新的以及与习惯性刺激存在差异的事物时。我们的各种感官都会产生定向反应，同时各种智能处理器会通过合作或竞争的方式在全局工作空间中展示它们对该新事物的认知分析方案，直到获得最佳的结果。全局工作空间可以看作信息共享的黑板系统，通过使用黑板，各个处理器试图传播全局性的信息，联合建立问题解决的办法。

全局工作空间通过竞争选出最突出、最相关、最重要和最紧迫的事件，它们的内容就成为意识的内容。然后，这些意识的内容被广播到全空间，启动行动的选择阶段。

注意是复杂的认知功能，这是人类行为的本质。注意是一个外部选择过程（声音、图像、气味……）或内部（思维）事件都必须保持一定水平的觉知。根据给定的语境情况下，选择性或集中注意力的选择在信息上应优先处理。选择性注意使你专注于一个项目，而明智地识别和区分不相关信息。CAM 采用兴趣度策略来实现注意选择。

动机是直接推动个体活动以达到一定目的的内在动力和主观原因，是个体活动的引发和维持的心理状态。在心智模型 CAM 中，动机模块的实现通过短时记忆系统完成。在 CAM 系统中，信念记忆存储智能体当前的信念，包含了动机知识。愿望是目标或者说是期望的最终状态。意图是智能体选择的需要现在执行的目标。目标/意图记忆模块存储当前的目标和意图信息。在 CAM 中，目标是由子目标组成的有向无环图，执行时分步处理。一个个子目标按照有向无环图所表示的路径完成，当所有的子目标都完成之后，总目标完成。对于一个动机执行系统来说，最关键的就是智能体内部的规划部分。通过规划，每个子目标通过一些列的动作来完成，从而，最终实现我们所希望看到的任务。规划主要处理内部的信息和系统新产生的动机。

元认知模块为智能体提供关于自己思维活动和学习活动的认知和监控，其核心是对认知的认知。元认知模块具有元认知知识、元认知自我调节控制和元认知体验的功能。元认知知识包括关于主体的知识，任务的知识以及策略的知识。元认知体验指的是对于自己认知过程的体验。在认知过程中，通过元认知自我调节控制，选择合适的策略，实现策略的使用，进程与目标的比较，策略的调整等等。

内省学习模块是通过检查和关注智能系统自身的知识处理和推理方式，从失败或低效中发现问题，形成修正自身的学习目标，由此改进自身处理问题的方法。在一般内省学习模型的基础上，采用本体技术构建知识库。内省学习系统中的一个重要问题就是失败的分类问题。失败的分类是诊断任务的基础，同时它为解释失败和构建修正学习目标提供重要的线索。失败分类需要考虑两个重要的因素，一个是失败分类的粒度，另一个是失败分类、解释失败及内省学习目标的关系。基于本体的知识库是将基于本体的知识表示方式同专家系统的知识库相结合，从而知识库具有概念化、形式化、语义明确、共享等优点。通过利用基于

本体的知识库方法解决内省学习中的失败分类问题,使得失败分类更加清晰,检索过程更加有效。

关于心智模型 CAM 的机器意识系统的详细内容,请参阅作者的著作《心智计算》[931]。

## 13.11 注意

注意是心理活动或意识在某一时刻所处的状态,表现为对一定对象的指向与集中。在大多数时候人们可以有意识地控制自己的注意方向。注意有两个明显的特点:指向性和集中性。注意的指向性是指人在每一瞬间的心理活动或意识选择了某个对象,而忽略了其余对象。在大千世界中,每时每刻都有大量的信息作用于我们,但是,我们无法对所有的信息都作出反应,只能把我们的意识指向其中一些事物。例如,你去商店买东西,你只注意到了你需要的东西,而忽略了其他的商品。所以,注意的指向性是指心理活动或意识在哪个方面上进行活动。指向性不同,人们从外界接收的信息也不同。当心理活动或意识指向某个对象的时候,它们会在这个对象上集中起来,即精神贯注,兴奋性提高。这就是注意的集中性。可以说注意的指向性是心理活动或意识朝向哪个对象,那么,集中性就是指心理活动或意识在一定方向上活动的强度或紧张程度。人在高度集中自己的注意时,注意指向的范围就缩小;指向的范围广泛而不集中时,整个强度就降低。人在注意高度集中时,除了对目标事物之外,对自己周围的其他事物就都会变得视而不见、听而不闻了。

### 13.11.1 注意的功能

注意具有选择功能、维持功能、调节功能等。我们周围的环境随时提供着大量的刺激,但这些信息对我们来说具有不同的意义。有的信息是重要的、有益的,但也有的信息于我们所从事的任务无关,甚至是一些有害的干扰信息,注意的第一个功能就是从大量的信息中选择重要的信息给出反应,同时排除掉有害的信息干扰,这就是选择功能。注意能够使人的心理活动或意识在一段时间内保持比较紧张的状态,就是要靠注意的维持功能。人只有在持续的紧张状态下,才能够对被选择的信息进行深入的加工与处理。注意的持续功能还体现在时间的延续上,对于复杂活动的顺利进行有重要意义。注意的调节功能不仅表现在稳定而持续的活动中,而且也表现在活动的变化中。当人们要从一种活动转到另一种活动时,注意体现了重要的调剂作用。有了注意的调节功能,才能实现活动的转变,也才能适应瞬息万变的环境。

#### 1. 定向控制

定向控制是指大脑把注意焦点导向感兴趣的地点、实现空间选择的能力。选择空间信息的方法有两种:第一种注意转移机制涉及眼睛的注视机制。受视野中的突出目标或个人意志的驱动,观察者的眼睛移到感兴趣的地点,并注视相应的目标。眼睛通过注视机制,使目标成像在视网膜的中央凹,从而获得较详细的目标信息。这种依靠眼动实现的定向控制和注意转移称为显式注意转移。第二种注意转移机制不涉及任何眼动或头动,它发生在两个大的跳动性眼动之间,以隐蔽的方式把注意力转向注视点之外的某个位置。这种注意转

移称为隐式注意转移。波斯纳(M I Posner)认为,隐式注意转移可能涉及三种注意操作:从当前的注意焦点解除注意(涉及大脑顶叶);把注意指针移向目标所在区域(由中脑区负责);在注意的指针处读取数据(丘脑枕核的功能)。人类具有隐式注意转移的能力。实验发现,当通过注意线索把注意点隐蔽地转向注视点之外的某个位置时,被试者不但提高了对该地点刺激的响应速度、降低了检测阈值,还增强了相应的头皮电活动。注意的定向性还说明,我们不能同时注意视野中的多个目标,只能一个一个地依次移动注意点,也就是说只能采用串行移动方式。但我们可以选择与视觉输入相应的加工尺度。注意点既可以精细聚焦,也可以散布在较宽的空间范围。在注意的认知模型中,把注意焦点比喻为可变焦距的聚光灯,就形象地反映了这种特性。

注意的定向选择性与注意系统的有限信息处理能力有关。被注意地点信息处理效率的增强以非注意地点信息被抑制为代价。

临床观察说明,对于大脑右顶叶受伤的病人,当注意线索呈现在右视野、而目标呈现在左视野时,定向控制能力损害严重,但在其他情况下,则损害不大。这表明从诱导线索地点解除注意的能力受损。由正常的被试者获得的 PET 数据也显示,当注意点从一个地方移动到另一个地方时,不管这种移动是由意志驱动的还是由外界刺激驱动的,血流明显增加的区域主要集中在左右上顶叶。这是唯一由注意转移激活的区域。在清醒猴子的顶叶所做的细胞记录也说明,顶叶神经元涉及注意的定向控制。PET 研究还显示,有选择地调制其他纹外皮层的解剖网络穿过丘脑枕核,滤掉干扰或增强目标的操作也在丘脑枕核中引起明显的效应。

PET 测量和临床观察还表明,大脑两半球的注意功能是不对称的。左右两个视野中的注意移动,可以引起右侧上顶叶的血流增强;而左侧上顶叶的血流增加只与右视野的注意移动有关。这个发现也许可以用来解释为什么大脑右半球的损伤比左半球的损伤引起更大的注意损害。然而,对于大脑正常的被试者,左右顶叶整合为一个单一的机制,因此隐蔽的注意也只涉及单一的中心。大脑的胼胝体对统一两半球的注意焦点起了关键作用。在视觉搜索任务中,对于正常的被试者,当同样数目的干扰目标分布在左右两视野时,并不比集中在单一视野时能更快地完成搜索任务。但对切除胼胝体的病人,当干扰目标分散到双视野时,病人搜索目标的速度比干扰目标集中到单一视野时要快 2 倍。这说明,在胼胝体损伤之后,左右半球的注意机制解除了相互联系。

### 2. 指导搜索

在视觉搜索任务中,注意的指导作用十分明显。一般来说,被试者发现目标的时间随干扰目标的数目而线性增加。然而,要找出某个特定目标,并不需要对所有目标进行搜索。有确凿的证据说明,搜索可以在目标的非位置特征指导下进行。这些特征包括颜色、形状、运动等。

实验说明,当注意颜色、形状或运动特征时,大脑额叶区的神经活动明显增强,而在大脑顶叶区并没有发现放大效应。这说明指导搜索是前注意系统的职责。

在前注意系统中,前扣带回的作用统称为“执行功能”。“执行”包括两层含义:首先,大脑组织内部正在发生的处理过程必须通告“执行者”;然后,“执行者”对整个系统实施注意控制。实验发现,该区域的神经活动随目标数目的增加而增强,随训练的次数而减少。这与

注意的认知理论相吻合。在解剖上,前扣带回具有联结后顶区和前额区的神经通路。前额皮层的侧区在保持过去事件的表象中起了关键作用,而前扣带回涉及目标的清晰觉察和控制。这些发现说明,前注意系统可能是意志注意的神经基础和大脑发布注意命令的中枢[165]。这种猜测也许是不无道理的,因为人类大脑的额叶区正是与我们制订计划有关的皮层区域,是心理活动的最高控制中心。

实验还发现,通过颜色、形状等特征进行选择和通过位置进行选择是同时发生的,相互干扰较少。因此有人猜测,前注意系统和后注意系统可能采用了类似时间共享或分时的策略。

#### 3. 保持警觉

警觉系统的功能是使大脑做好准备和保持警觉,以便快速处理具有最高优先权的信号。保持警觉与注意密切相关,它涉及注意的一个子网络。

正电子断层扫描显示,当要求被试保持警觉状态时,大脑的右侧额叶区血流增强;而当该区域受损伤时,人类则丧失保持警觉的能力。这说明,保持警觉状态涉及一个位于大脑右侧的注意子系统。

#### 4. 抑制-增强效应

大的活动可通过注意的调控有选择地抑制或增强。注意的抑制-增强效应是三个注意子网络协同作用的结果。

当视野中存在大量的干扰目标时,大脑如何找到正确的地点从而完成目标检测呢?实验说明,大脑是通过有选择地放大或抑制各脑区的神经活动,完成目标检测的。PET 测量显示,当指示被试者注意刺激的某种属性时,专门负责处理该属性的大脑区域就被有选择地增强。这种效应在视觉系统纹外皮层尤其明显。实验发现,尽管被试者视野中的刺激图像完全相同,但不同的指导语即可引起不同脑区的活动增强。通过指导语让被试者注意视觉刺激的某种属性(如运动速度)和由被试者直接观察具有该属性的物理刺激(如运动目标)可以引起相同脑区的活动增强。一般说来,任何一个脑区域都可以通过注意的作用而增强。

用记录人脑头皮电位的方法同样测量到了注意的增强效应。在某些搜索任务中,仅凭朝向或形状等单一特征是不能发现目标的;要完成搜索任务,必须把两个以上的特征在同一地点结合起来。这时,被试需要进行串行搜索,把注意点从一个地方移动到另一个地方。如果这时不允许眼动,那就只能依靠隐蔽的注意转移。在实验中,隐蔽的注意转移是由呈现在不同地点的诱导线索驱动的。实验发现,当注意线索呈现在目标位置时,把注意焦点隐蔽地转移到注意线索出现的位置,就会引起被试者后顶区头皮电位的增强。这种增强效应发生的地点与正电子断层扫描研究中脑血流增强的皮层区域是完全相同的。

### 13.11.2 选择性注意

如果你参加鸡尾酒会或在嘈杂的饭店,有三种办法可以帮助你关注你想听的人(注意目标)的说话信息:一是目标话语与众不同的感觉特点(例如,高与低的音高、音步、节律);二是音强(响亮度);三是音源位置。注意目标说话者声音的物理特性之后,你就可以避免被该位置非目标说话者说话信息的语义内容所干扰。显然,目标的音强也有帮助。另外,你可

能凭直觉用某个策略来定位声音，这样就使得双耳同听任务变为双耳分听任务：你一只耳朵转向听目标说话者，而另一只耳朵避开目标说话者。

选择性注意机制常见有过滤器模型、衰减模型、反应选择模型、能量分配模型等。

### 1. 过滤器模型

该模型最早由英国著名心理学家布罗德贝特(D Broadbent)于 1958 年提出[103]，是关于注意的一个较早的理论模型。该模型后来被韦尔弗特(Welford)称为单通道模型。过滤器模型认为，来自外界的信息是大量的，但人的神经系统高级中枢的加工能力极其有限，于是出现瓶颈。为了避免系统超载，需要过滤器加调节，选择一些信息进入高级分析阶段，而其余信息可能暂存于某种记忆之中，然后迅速衰退。

注意的作用就像过滤器一样，通过它的过滤，使输入通道的一些信息能够通过内部容量有限的通过进入高级中枢过程，而其他的信息则被过滤掉，参见图 13.3。因此，注意的工作是以全或无的方式进行的。

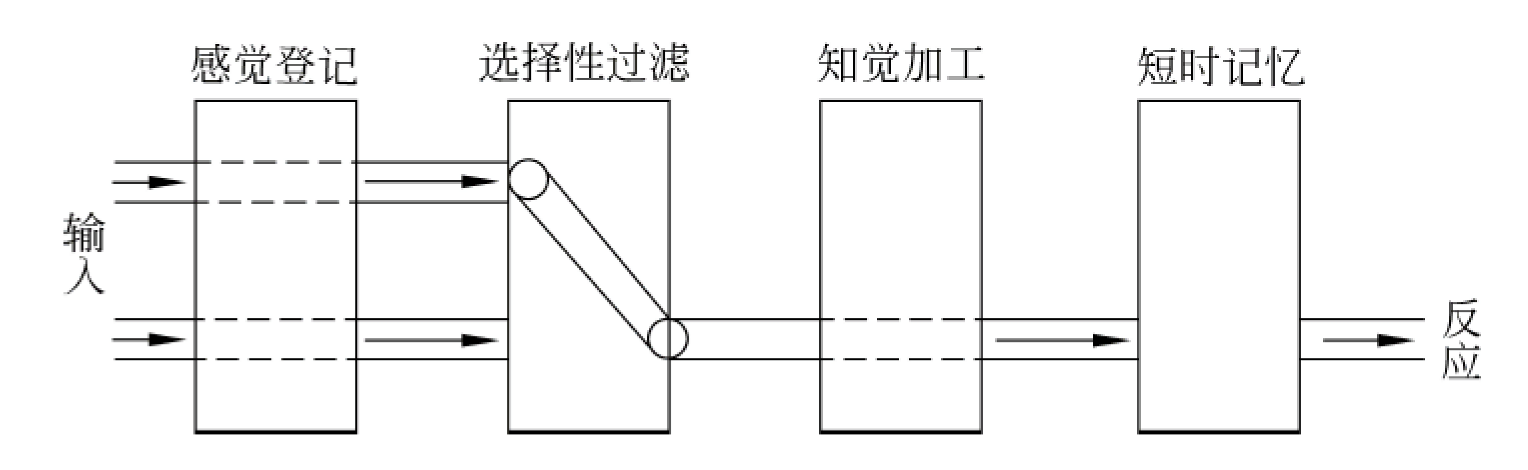



图 13.3　过滤器模型

在鸡尾酒会或其他的聚会上，你正在和几个人专注地交谈。这时，你对其他人的交谈是不能识别的，但如果外面有人提起你的名字，你可能会注意到，而与你交谈的其他人却不一定注意到。通过耳麦给被试两耳同时放音，每只耳朵所接收的刺激信息是不一样的。通过实验考察被试反应信息与双耳接收信息的关系，可以了解被试注意的特点。实验结果发现被试在这样的实验中会根据材料的特点重现刺激信息，如 6、2、9；DEAR AUNT JANE。

对于这种实验结果，存在如下两种看法。

(1) 过滤器通道可以快速转移。

(2) 注意的工作机制不是单通道模型和以全或无的方式进行工作，过滤器模型不正确。

### 2. 衰减模型

衰减模型是美国心理学家特瑞斯曼(A Treisman)于 1960 年在修正过滤模型的基础上提出来的[762]。该理论认为过滤器并不是按“全或无”的方式工作。接受信息的通道不是单通道，而是多通道。该模型是特瑞斯曼根据追随实验中对非追随耳信息的信息也可以得到加工的实验结果，对过滤器模型加以改进的结果。

衰减模型认为，注意的工作方式不是以全或无的单通道方式，而是多通道方式工作，但是在多通道的信息加工中，每个通道的信息得到加工程度是不一样的。如图 13.4 所示。追随耳的信息加工方式如过滤器模型所述，而非追随耳的信息也有可能通过过滤器而被加工。只是非追随耳的信号在通过过滤器信号被衰减，故以虚线表示，在意义分析的过程中，有可

能被过滤掉,也有可能因为其他一些因素而被加强(如自己名字)。

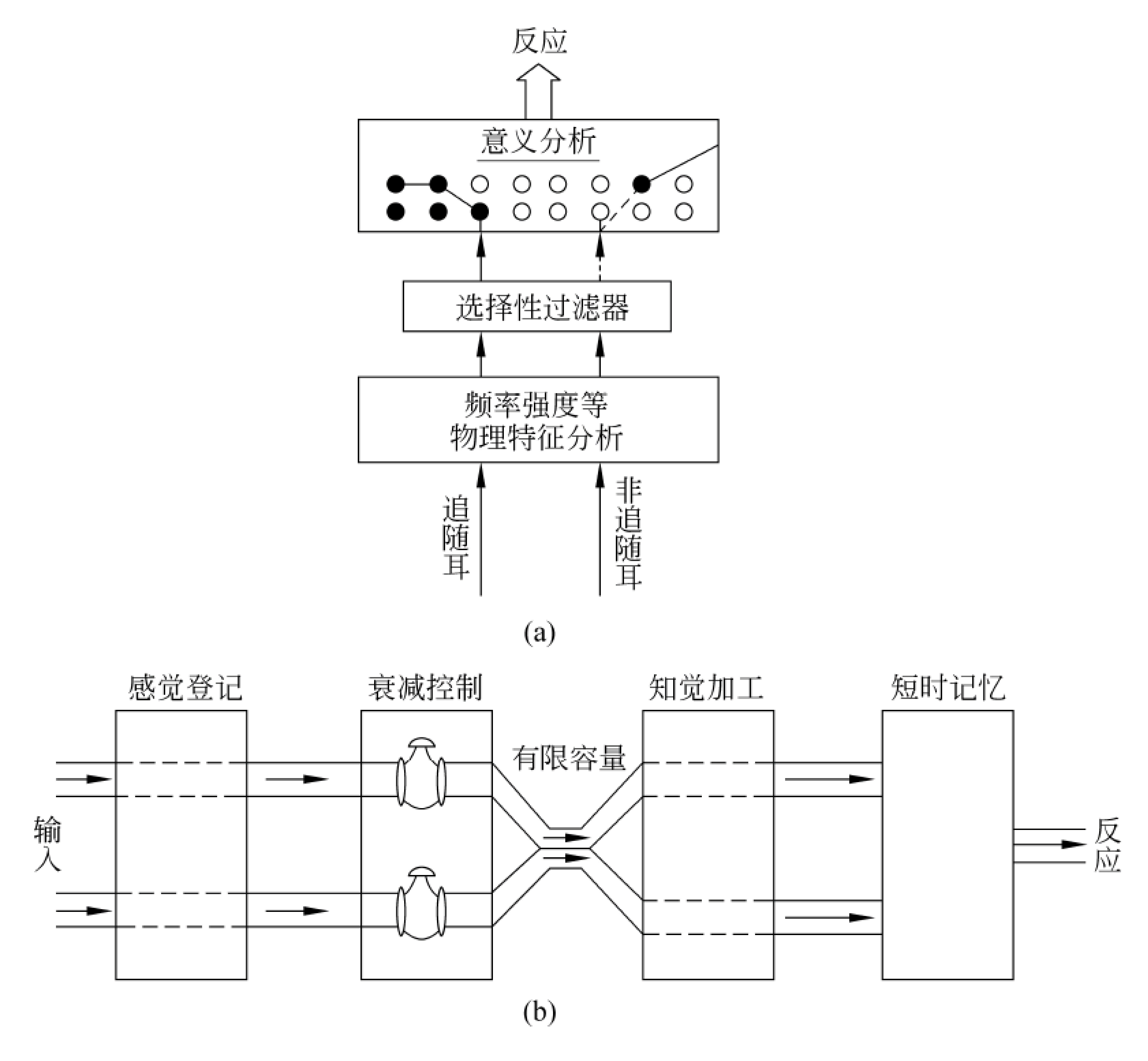



图 13.4 衰减模型

按照特瑞斯曼的理论看,选择性注意涉及三个阶段。在第一个阶段我们事先注意并分析了刺激的物理特性,如音量(声音强度)、音高(声波的"频率")等。这个预先注意加工在前来的感觉刺激上并行(同时)加工。对于具有目标刺激特性的刺激,我们把信号传到下一个阶段,而对于没有目标刺激特性的刺激,我们把它仅仅当作弱化刺激。在第二阶段,我们要分析给定刺激是否有类似言语或音乐的模式。对那些具有目标模式的刺激,我们把信号传到下一个阶段,对于没有目标模式的刺激,我们也把它仅仅当作弱化刺激。在第三个阶段我们把注意集中在达到该阶段的刺激上,按序列评价前来的信息,赋予所选择的刺激信息适当的意义。

衰减模型对进入高级分析水平的信息引入一个重要的概念——阈限。认为已储存在大脑中的信息在高级分析水平中的兴奋阈限各不相同,从而影响着过滤器的选择。

### 3. 反应选择模型

1963 年,德意志(J Deutsch)提出反应选择模型[186]。该模型(见图 13.5)认为,注意并不在于选择知觉刺激,而在于选择对刺激的反应。该理论认为,感觉器官感受到的所有刺激都会进入高级分析过程。中枢则根据一定的法则进行加工,对重要的信息才作出反应,而不重要的信息可能很快被新的内容冲掉。

与特瑞斯曼的衰减理论不同的之处仅仅只是把信号阻断过滤的位置放在一些识别

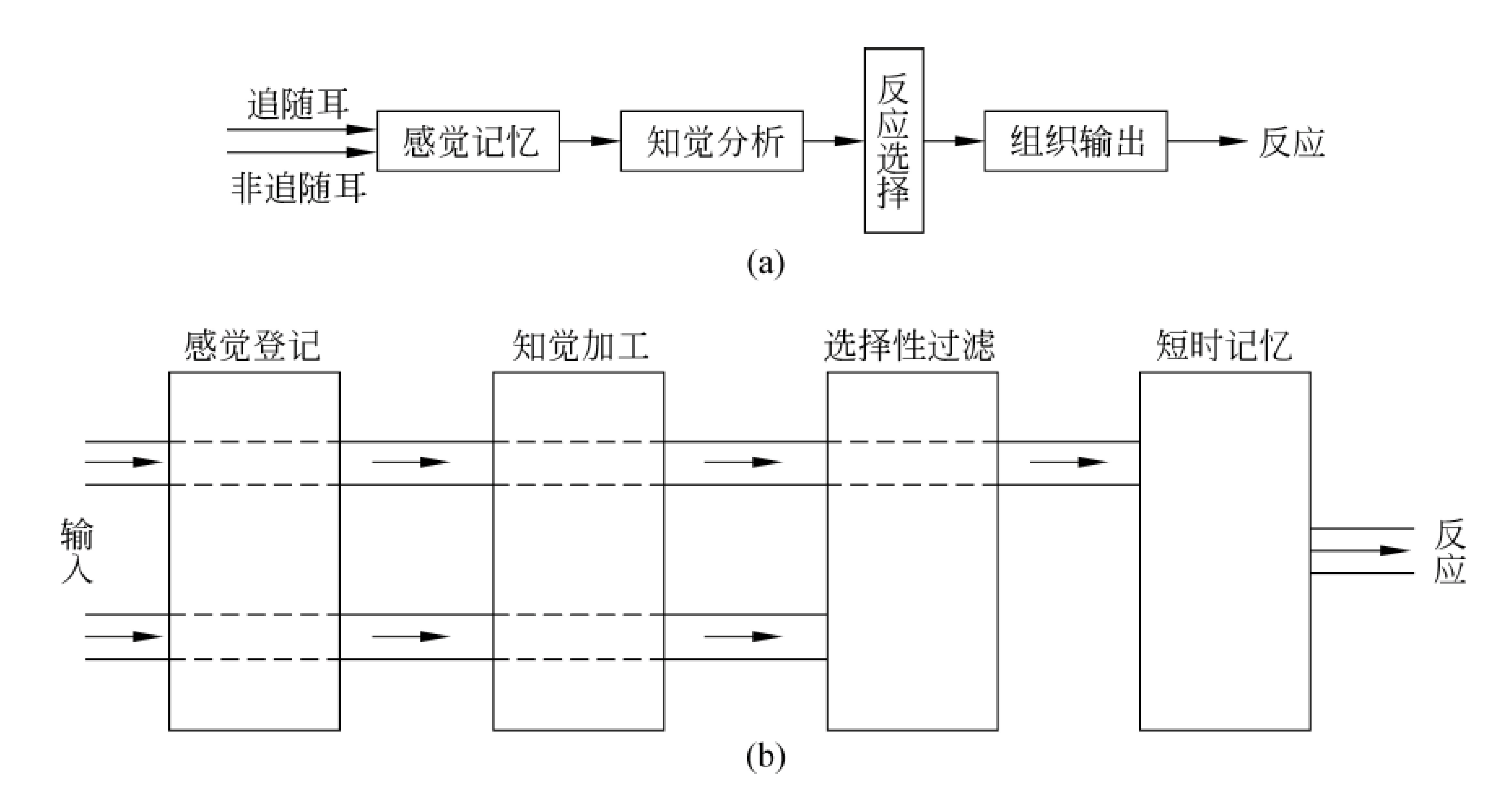



图 13.5　注意的反应选择模型

刺激的意义所需要的知觉加工之后，而不是之前。这个后过滤使得人们能够识别进入到非注意耳的信息。比如，听到自己的名字或输入的翻译（对双语者而言）。如果信息没有在认知觉上触动某根弦，那么人们就会在过滤机制中把它抛除掉。如果有，如听到自己的名字，人们就会注意它。需要指出的是，后过滤机制和前过滤机制的支持者都提出，存在只允许单一的信息源通过的注意瓶颈，这两个理论模式的区别仅仅是瓶颈的假定位置不同。

### 4. 资源分配模型

1973 年，卡尼曼（D Kahneman）提出资源分配模型[357]。资源分配模型认为，注意是人能用于执行任务的数量有限的能量或资源。人在活动中可以得到的资源和唤醒是连接在一起的，其唤醒水平受情绪、药物、肌紧张等因素的影响，所产生的资源通过一个分配方案被分配到不同的可能活动之中，最后形成各种反应。图 13.6 给出了资源分配模型的示意图。

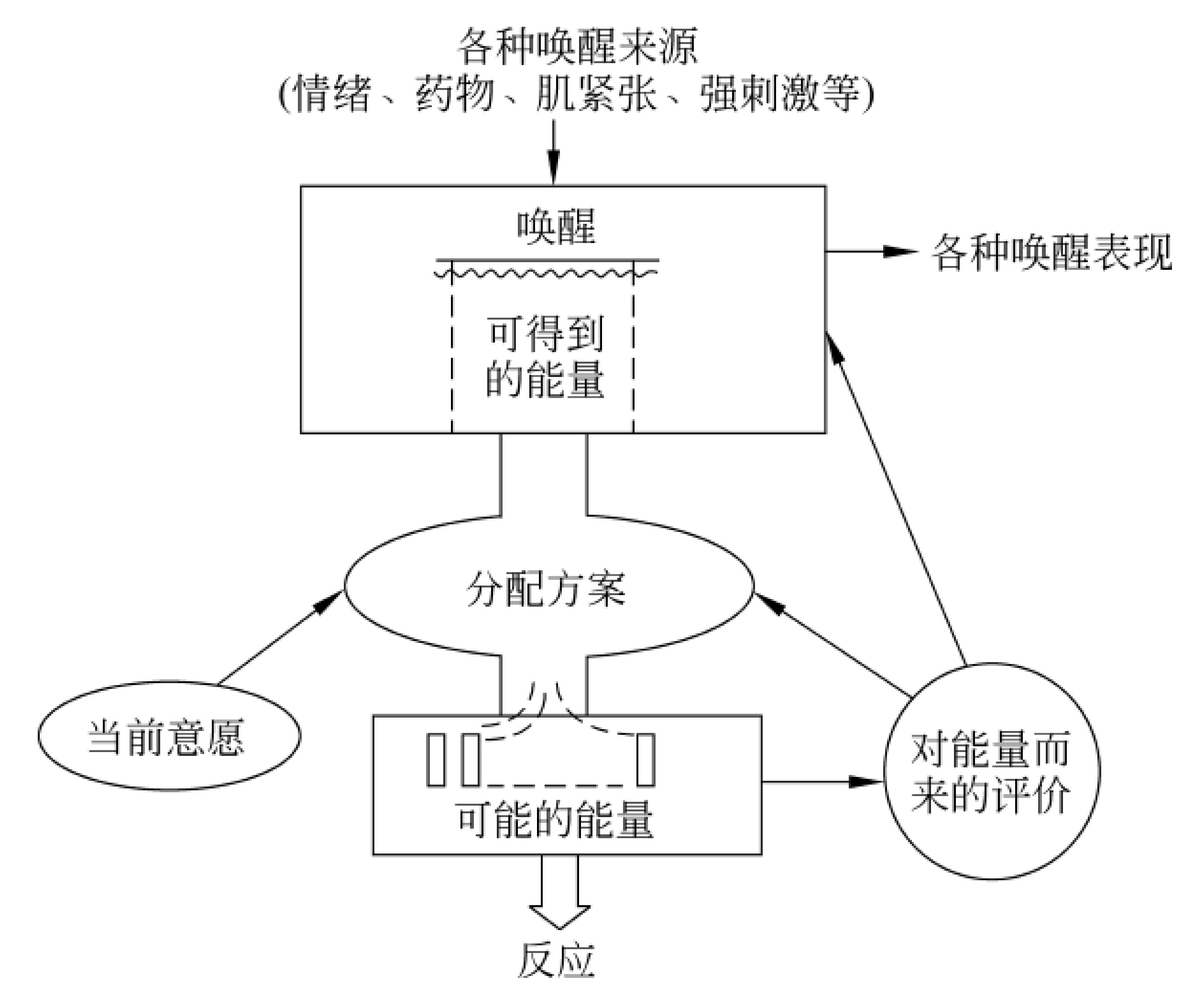



图 13.6　资源分配模型

### 5. 聚光灯模型

20 世纪 80 年代初期,特雷斯曼提出的特征整合模型把注意和知觉加工的内部过程紧密地结合起来,并用"聚光灯"形象地比喻注意的空间选择性[763],受到认知心理学家、神经生理学家以及神经计算学家的高度评价。根据这一模型,视觉处理过程被分为两个相互联系的阶段,即预注意和集中注意阶段。前者对视觉刺激的颜色、朝向和运动等简单特征进行快速、自动的并行加工,各种特征在大脑内被分别编码,产生相应的"特征地图"。特征中的各个特征构成预注意的表象。预注意加工是一个"自下而上"的信息处理过程,并不需要集中注意。特征地图中的各个特征在位置上是不确定的,要获得物体知觉就需要依靠集中注意,通过"聚光灯"对"位置地图"进行扫描,把属于被搜索目标的各个特征有机地整合在一起,实现特征的动态组装。这是速度较慢的串行过程。因此,当目标与干扰目标的差别仅仅表现为单个特征时,目标可以立刻从视野中"跳出",完成目标检测,搜索时间不受干扰目标数目的影响;而当目标与干扰项的差别表现为多个特征的结合时,就需要运用集中注意,对各个目标位置顺序扫描。这时,搜索时间随干扰项的数目线性增加。

最近的实验证据表明,根据较复杂的属性,如三维表面,也可以实现注意分配。特别令人感兴趣的是,特征整合模型中的特征组装构想已在视觉神经生理研究中得到部分证实。大脑中并不存在与我们所看到的物体一一对应的众多皮层神经单元,相反,物体进入视觉系统后,被分解为不同的特征或属性,如颜色、朝向、运动等。这些特征由不同的视觉通道和皮层区域分别处理。如何用不同区域神经元的发放统一表征同一目标的所有属性,或能否找到进行相关发放的神经元的共同标记,即所谓的"组装问题"。电生理实验表明,集中注意可以引起与被注意事件相关的神经元的同步发放。这种同步发放通常表现为 40 周左右的同步振荡。这一发现为注意的特征整合模型提供了神经生理证据。然而,注意的"聚光灯"如何从一个位置移向另一个位置,这仍然是一个有待解决的问题。

注意的"聚光灯"模型是一个具有广泛影响的认知心理学模型。然而,它并没有被神经生理学家普遍接受。有些理论甚至对注意"聚光灯"的客观存在也提出了疑义。例如,注意的竞争理论就认为,注意机制并不是视野中快速扫描的"聚光灯",而是大脑的多种神经机制为解决物体竞争和行为控制而产生的涌现特性。

尽管目前还没有被普遍承认的注意理论,但注意涉及一个瓶颈问题则是大家所公认的,这就是注意的选择性问题。其基本的思想就是初级信息处理大体上是快速、并行的;而在信息处理的某个或某些阶段就需要一个瓶颈去控制信息处理,一个时刻只能处理一件事,而其他位于非注意点的事件则被暂时抑制掉。然后再快速移动到另一件事情,依此类推。这是一个串行过程,需要较长的时间去完成[164]。

## 13.11.3 注意分配

人在进行两种或多种活动时能把注意同时指向不同对象的现象。这种现象在生活中到处可见。例如,教师一边讲课,一边还能观察学生听讲的情况;汽车司机一边开车,一边打电话,还要注意周围环境的变化等。

从理论上讲,大脑在同一时刻内信息加工的容量有限,注意是对于其中一部分信息的集中。集中注意的加工要求全部心理活动参加,有很高的紧张性,因此注意不可能同时指向两

个不同方向。即注意分配与注意集中是相矛盾的。关于注意不能分配或难于分配，自古以来就有很多实验证明了。近代用复合器做的实验发现，不同种类的刺激严格地同时作用于两个感官的时候，注意分配相当困难。复合器的构造是，它有一个划分为 100°的圆刻度盘，盘上有一根迅速转动的指针，当指针经过某一刻度时，就会响起铃声。被试的任务是在响铃时，说出指针所指的度数。结果表明，被试通常不能说出铃响时指针所指的准确度数，他所说出的总是早于或晚于铃响时的度数。这说明他的注意首先指向于一个刺激物(铃声或指针的位置)，而稍迟一些时间才能指向另一刺激物。还有研究表明，严格地同时给予两耳以不同的信息，要被试同时感受它们，即进行双耳听辨的实验，结果也难以准确反映。

研究表明，注意分配有不同的水平，它取决于同时并进的几种活动的性质、复杂程度以及人对活动的熟悉或熟练程度等条件。当同时进行的几种活动复杂或难度越大时，注意分配就越困难。在智力和运动两种活动同时进行时，智力活动的效率比运动活动的效率有更大程度的降低。同时进行两种智力活动，则注意分配的困难更大一些。在影响注意分配的各种因素中，对活动的熟练程度起作用最大。要想能够很好地分配注意，首先是在同时进行的两种活动中，必须有一种活动达到了相对“自动化”的程度，即不再需要更多的注意，这样人就能把注意集中在比较生疏的活动上。其次，使同时进行的几种活动之间建立一定的联系，或通过训练使复杂的活动形成一定的反应系统，这样注意分配也就比较容易了。

注意的分配与转移是密切联系的。所谓注意转移，是人能根据一定的目的，主动地把注意从一个对象转移到另一个对象上。为了顺利完成某项复杂的活动，注意中心在不同对象间的迅速往返转移，就构成了注意分配现象。如果这种往返转移的速度要求太快，超过了注意转移的能力，注意的分配也就难以实现。复合器实验和双耳听辨实验的结果就是例证。

### 13.11.4 注意系统

根据已有的研究结果，波斯纳把注意网络分为 3 个子系统[593]：前注意系统、后注意系统和警觉系统。前注意系统主要涉及额叶皮层、前扣带回和基底神经节。后注意系统主要包括上顶皮层、丘脑枕核和上丘。警觉系统则主要涉及位于大脑右侧额叶区的蓝斑去甲肾上腺素到皮层的输入。这三个子系统的功能可以分别概括为定向控制、指导搜索和保持警觉。前注意系统在需要意识的任务中激活得越来越多，其间被试必须注意单词的词义。这个系统也关系到“行动注意”，其中被试在选择性的行动过程中计划或选择一项行动。相反，后注意系统关系到皮质顶叶。顶叶是丘脑的一部分，而有些中脑区域与眼动相关。进行视觉空间注意期间，该系统高度激活，其中被试必须脱离或转换注意。进行视觉、听觉、运动或高级任务时皮质的视觉、听觉、运动或联想区域的神经活动也与注意相关。前注意系统和后注意系统看来能增强各个不同任务的注意。这表明，它们用来调节特定任务的相关皮质区域的激活。

注意系统的活动是否是注意项目激活度的增强、非注意项目的抑制或压制激活所导致的结果，或者二者都有？波斯纳等对此做出了回答：要看情况而定。具体来讲，根据特定任务以及所考察的大脑区域不同，这三种情况都有可能。所要研究的任务决定了执行什么任务时大脑的什么区域发生什么样的加工。为了勾画各种任务所涉及的大脑区域，认知神经心理学家经常用正电子发射层析摄影术(PET)，它可以勾画出区域性的皮质血液流动。

# >>第 14 章
# 认知结构

认知结构是指认知活动的组织形态和操作方式，包含了在认知活动中的组成成分及成分之间的相互作用等一系列的操作过程，即心理活动的机制。认知结构理论以认知结构为研究核心，强调认知结构建构的性质、认知结构与学习的互动关系。

## 14.1 概述

智能科学探索智能的机理，要研究认知结构的组织形态和操作方式。纵观认知结构的理论发展，主要有皮亚杰的图式理论、格式塔的顿悟理论、托尔曼的认知地图理论、布鲁纳的归类理论、奥苏伯尔的认知同化理论等。

(1) 皮亚杰的图式理论。

皮亚杰认为图式是主体的认知结构，图式的建构过程是在同化和顺应两种共同作用中完成的。皮亚杰从主-客体关系入手，认为主体认知结构根源是主体动作基础上的内化建构。动作是认知结构的根源。认知结构经历了从感知运动图式→表象图式→直觉思维图式→运演思维图式的发展。

(2) 格式塔的顿悟理论。

韦特海墨(Wertheime)认为学习是知觉的重新组织，是构造一种“完形”，即格式塔。学习过程中问题的解决，都是由于对环境中事物关系的理解而构成一种“完形”所实现的，学习的成功与实现完全是由顿悟决定的[855]。

(3) 托尔曼的认知地图理论。

托尔曼坚持学习的符号-格式塔模式，有机体习得的是关于周围环境、目标位置以及达到目标的手段和途径的知识，也就是形成认知地图的过程，而不是简单、机械的反应。所谓认知地图，是关于某一局部环境的综合表象，它不仅包括事件的简单顺序，而且包括方向、距离甚至时间关系等。在认知地图的不断改造重组的过程中，有机体不断地获得关于环境的知识，形成综合的表象，达到目标符号。

(4) 布鲁纳的归类理论。

布鲁纳认为认知结构就是归类后的类别(概念、知识经验等)按层次水平的高低组成的编码系统。学习的实质就在于主动地形成认知结构。布鲁纳提出了认知结构发展的三个阶段：动作表征、映象表征和符号表征。儿童最初的认知结构，就是动作表征，他们“从动作中认知”，即他们的认知多数是通过行为而产生的。

(5) 奥苏伯尔的认知同化理论。

奥苏伯尔认为，所谓认知结构，是指个体具有知识的数量、清晰度和组织方式，它由事

实、概念、命题、理论等构成。它是个体对世界的知觉、理解和思考的方式。关于认知结构中的内容，奥苏伯尔做了创造性的分析，称为认知结构变量，它是指个体认知结构中的概念或观念有其组织方面的特征。

认知结构理论认为存在于人头脑中的认知结构始终处于变动与建构之中，学习过程就是认知结构不断变化和重组的过程，其中，环境和学习者的个体特征是决定性因素。皮亚杰用同化、顺应、平衡等过程表征认知结构建构的机制，强调了外在整体环境的重要性，认为环境为学习者提供的丰富、良好的多重刺激是促使认知结构完善和发生变化的根本条件。现代认知心理学家奈瑟认为，认知过程是建构性质的，它包括两个过程：个体对外界刺激产生反应的过程和学习者有意识地控制、转换和建构观念和映象的过程。认知结构就是在外在刺激和学习者个体特征相结合的情况下进行具有渐进性的自我建构的过程。

为了研究认知结构的无矛盾性问题，需要以逻辑系统为基础。在莱布尼茨的思想中，数理逻辑、数学和计算机三者均出于一个统一的目的，即思维过程的演算化、计算化，以至在计算机上实现。早在 20 世纪 30 年代，数理逻辑将推理化为一些简单机械的动作，提出了图灵机这一计算机的抽象模型，并证明了存在通用图灵机，这正是 20 世纪 40 年代出现的存储程序计算机（即冯・诺依曼计算机）的理论原型。

符号逻辑倡导的形式化方法已广泛渗入到各个领域。程序逻辑、算法逻辑、动态逻辑、时态逻辑在形式化方法中有许多应用。在儿童心理学的研究中，皮亚杰改造了数理逻辑并用来描述儿童不同智力水平的认知结构。这种用来刻画儿童不同智力阶段认知结构的逻辑在目的、特点和作用等方面都不同于经典的数理逻辑，形成心理逻辑系统。

## 14.2　谓词演算

在命题演算中，每个原子符号（$P$、$Q$ 等）表示某种复杂度的一个命题。没有办法访问断言的各个部分，谓词演算提供了解决这一问题的能力。例如，不再是让一个命题符号 $P$ 表示整个句子“星期二下了雨”，而是创建一个谓词 weather 来描述日子和天气的关系：weather(Tuesday,rain)。通过推理规则我们可以操纵谓词演算表达式，访问它的每个组成成分，而且推理出新的语句。

谓词演算还允许表达式中含有变量。变量使我们可以建立关于实体类的通用断言。例如，我们可以声明对于所有的 $X$ 值，其中 $X$ 是某一周的一天，陈述 weather($X$，rain)是真的；也就是说这一周一直下雨。全称量词 ∀ 指出其后的语句对于变量的所有值为真。例子中，$\forall X$ likes($X$, ice_cream)对于 $X$ 定义域中所有值为真。存在量词∃指出至少对于定义域中的一个值语句为真。

谓词演算的语义为逻辑推理的正规理论提供了基础。从一系列真实断言推理出新的正确表达式的能力是谓词演算的一个重要功能。这些新的表达式正确的标准是它们与对原始表达式集合的所有以前解释是一致的。

研究由个体、函数及关系构成的命题以及由这些命题经使用量词和命题联结词构成的更复杂的命题和这类命题之间的推理关系的逻辑。在一阶逻辑中，量词仅作用于个体变元。一阶逻辑是数理逻辑中发展得最为成熟的部分。在为数学的语言和推理建立形式系统的过程中，它处于核心地位，又称谓词逻辑。

传统逻辑主要是指古代希腊的亚里士多德逻辑,在中世纪被认为是金科玉律,完善无缺,不容许有任何更改。但是到了19世纪,人们觉得它有很多缺点,需要改革。传统逻辑仅限于讨论主宾式语句和三段论形式的推理,并且缺乏对于量词的研究。德·摩根(A De Morgan)研究和发展了关系逻辑,提出了论域的概念。哈密顿(W Hamilton)对量词进行了研究。弗雷格(G Frege)于1897年建立了第一个谓词逻辑的形式系统。罗素(B Russell)和怀特海(A Whitehead)于1910年在他们的数学名著《数学原理》中总结了前一段的成果,建立了一个完全的谓词演算。1930年,哥德尔(K Godel)证明了谓词演算的完备性。

在命题逻辑中,不进一步分析原子命题的内部结构,原子命题被看作是不可分割的最小单位,因而不能包括某些正确的推理。例如以下的推理:

每个有理数都是实数,1是有理数,所以,1是实数。

在一阶逻辑中描述一个数学理论,首先会涉及这个理论所讨论的对象、定义在这些对象上的函数以及这些对象之间的关系。数学理论所讨论的对象称为个体,由个体组成的非空集合称为论域或个体域。相等关系是经常需要用到的,在一阶逻辑中用一个特殊的符号,即等号“=”表示它。

为了表达每个个体都有某性质,在一阶逻辑中引进了全称量词任意取。为了表达至少有一个个体有某性质,在一阶逻辑中引进了存在量词。例如,设论域是整数集,$N(x)$表示$x$是自然数。$\forall xN(x)$表示命题“每个整数都是自然数”,这是一个假命题。$\exists xN(x)$表示命题“至少有一个整数是自然数”,这是一个真命题。

一阶逻辑使用的形式语言称为一阶语言,它的符号包括以下几类。

(1) 个体变元$x,y,z,\cdots$,简称为变元。

(2) 函数符号$f,g,h,\cdots$;个体常元$a,b,c,\cdots$,谓词符号$P,Q,R,\cdots$,其中包括二元谓词符号“=”。

(3) 命题连接词$\neg$,$\rightarrow$和全称量词$\forall$。

通常称函数符号、个体常元、除等号“=”之外的谓词符号为非逻辑符号,而称其余符号为逻辑符号。不同的一阶语言有不同的非逻辑符号,而所有一阶语言的逻辑符号都是相同的。因为$\wedge$,$\vee$可用$\neg$,$\rightarrow$定义,所以可取$\neg$,$\rightarrow$为基本联结词(参见命题逻辑)。因为存在量词$\exists$可用$\forall$和$\neg$定义,$\exists xN(x)$与$\neg\forall x\neg N(x)$表达的命题之真假意义相同,所以仅取$\forall$为基本量词。

项的形成规则如下。

(1) 变元和个体常元是项。

(2) 若$f$是$n$元函数符号,$x_1,\cdots,x_n$是项,则$f(x_1,\cdots,x_n)$是项。

(3) 每个项都可以通过有穷次应用(1)和(2)获得。

公式的形成规则如下。

(1) 若$P$是$n$元谓词符号,$x_1,\cdots,x_n$是项,则$P(x_1,\cdots,x_n)$是公式,也称为原子公式。

(2) 若$A$是公式,则$\neg A$是公式。

(3) 若$A,B$是公式,则$(A\rightarrow B)$是公式。

(4) 若$A$是公式,$x$是变元,则$\forall xA$是公式。

(5) 每个公式都可以通过有穷次应用(1)~(4)获得。

如果变元$x$出现在公式$A$中形如$\forall x B$的部分,则称$x$在$A$中的这次出现为约束出

现，否则称为自由出现。例如，在公式 $P(x)\to\forall xQ(x)$ 中，第一个 $x$ 是自由出现，后两个 $x$ 是约束出现。如果变元 $x$ 在公式 $A$ 中有自由出现，则称 $x$ 为 $A$ 的自由变元。没有自由变元的公式称为闭公式。常用 $[t]$ 表示将公式 $A$ 中 $x$ 的所有自由出现代之以项 $t$ 所得到的公式。如果 $[t]$ 和 $A$ 中的变元的约束出现数相同，则称 $t$ 对 $A$ 中的 $x$ 是可代入的。例如 $f(z)$ 对于 $P(x)\to\forall y\,Q(x,y)$ 中的 $x$ 是可代入的，而 $f(y)$ 对于 $P(x)\to\forall y\,Q(x,y)$ 中的 $x$ 不是可代入的。

设 $L$ 是一个一阶语言。可指定一个非空集合为论域，将等号"＝"解释为论域上的相等关系，为 $L$ 的每个个体常元指定论域中的一个个体，为 $L$ 的每个 $n$ 元函数符号指定论域上的一个 $n$ 元函数，为 $L$ 的每个非逻辑的 $n$ 元谓词符号指定论域上的一个 $n$ 元关系，就得到 $L$ 的一个结构。闭公式在一个结构中的解释是一个命题。例如，指定论域为正整数集，个体常元 $a$ 解释为 2，$P(x)$ 解释为 $x$ 是素数，$L(x,y)$ 解释为 $x\leqslant y$，则闭公式 $P(a)\wedge\forall y(P(y)\to L(a,y))$ 就解释为真命题：2 是最小的素数。有自由变元的公式在一个结构中的解释还不是一个命题，因为自由变元值不确定。例如，在上面所给的结构中，如果给变元 $z$ 赋值 2，则公式 $P(z)\wedge\forall y(P(y)\to L(z,y))$ 解释为真命题；如果给变元 $z$ 赋值 3，则该公式解释为假命题。设给定一个结构 U，如果给自由变元赋予论域中任何个体，公式 $A$ 都被解释为真命题，则称 $A$ 在 U 中有效，记为 $\mathrm{U}\models A$。如果公式 $A$ 在每个结构中有效，就称 $A$ 为逻辑有效式或永真式，记为 $\models A$。例如，公式 $\forall xP(x)\to P(y)$ 是逻辑有效式，而 $P(y)\to\forall xP(x)$ 不是逻辑有效式。

谓词演算把逻辑有效式组成了一个完全形式化的公理系统。在谓词演算中，取某些逻辑有效式为公理，并规定了一些推理规则，以推导出所有的逻辑有效式。人们给出了许多等价的谓词演算系统。这里介绍其中一个。取以下 7 种形式的公式为公理。

(1) $A\to(B\to A)$。

(2) $(A\to(B\to C))\to((A\to B)\to(A\to C))$。

(3) $(\neg B\to\neg A)\to(A\to B)$。

(4) $x=x$。

(5) $x=y\to(A\to[y])$，其中 $y$ 对于 $A$ 中的 $x$ 可代入。

(6) $\forall x\,A\to[t]$，其中项 $t$ 对于 $A$ 中的 $x$ 可代入。

(7) $\forall x(A\to B)\to(A\to\forall xB)$，其中 $x$ 不是 $A$ 的自由变元。

推理规则有以下两条。

(1) 分离规则：由前提 $A$ 和 $A\to B$ 推出结论 $B$。

(2) 概括规则：由前提 $A$ 推出结论 $\forall xA$。

谓词演算的定理是这样定义的。

(1) 每个公理都是定理。

(2) 如果 $A$ 和 $A\to B$ 是定理，则 $B$ 是定理。

(3) 如果 $A$ 是定理，则 $\forall xA$ 是定理。

(4) 每个定理都可以通过有穷次应用(1)～(3)获得。

若公式 $A$ 是定理，则记为 $\vdash A$。

谓词演算的公理都是逻辑有效式。推理规则保证，当前提都在某结构中有效时，结论也在该结构中有效。因此，谓词演算的定理都是逻辑有效式。这个性质称为谓词演算的可靠

性。反之,每个逻辑有效式都是谓词演算的定理。这是谓词演算的完备性,是由哥德尔于1930年证明的。

公式是不是逻辑有效式是半可判定的,不是可判定的(参见可计算性理论)。可以用一阶逻辑刻画数学理论,常把谓词演算的公理称为逻辑公理。除了逻辑公理之外,数学理论还需要非逻辑公理,它们刻画了该数学理论的研究对象的共同性质。这样的数学理论称为一阶理论。如果一阶理论 T 的每个非逻辑公理都在结构 U 中有效,就称 U 是 T 的模型。下面举出几个常见的一阶理论。

全序理论只有唯一的非逻辑符号,即二元谓词符号≤。它有4个非逻辑公理:

(1) $x \leqslant x$。

(2) $(x \leqslant y \wedge y \leqslant x) \rightarrow x = y$。

(3) $(x \leqslant y \wedge y \leqslant z) \rightarrow x \leqslant z$。

(4) $x \leqslant y \vee y \leqslant x$。

每个全序结构都是它的模型。

群论有两个非逻辑符号:个体常元 $e$ 和二元函数符号·。它有3个非逻辑公理:

(1) $\forall x \forall y \forall z((x \cdot y) \cdot z = x \cdot (y \cdot z))$。

(2) $\forall x(x \cdot e = x \wedge e \cdot x = x)$。

(3) $\forall x \exists y(x \cdot y = e \wedge y \cdot x = e)$。

每个群都是它的模型。

初等数论的非逻辑符号有:个体常元0,一元函数符号 $S$,两个二元函数符号+和·。它的非逻辑公理为:

(1) $\forall x \neg (S(x) = 0)$。

(2) $\forall x \forall y(S(x) = S(y) \rightarrow x = y)$。

(3) $\forall x(x + 0 = x)$。

(4) $\forall x \forall y(x + S(y) = S(x + y))$。

(5) $\forall x(x \cdot 0 = 0)$。

(6) $\forall x \forall y(x \cdot S(y) = (x \cdot y) + x)$。

(7) $A_x[0] \wedge (\forall x(A \rightarrow A_x S(x)])) \rightarrow \forall x A$。

自然数系统是它的一个模型。

一阶逻辑有很强的表达能力,用一阶理论能够刻画许多数学结构。但是,一阶逻辑不是万能的,其表达能力受到一定的限制。如果一阶理论 T 的模型的类恰好是结构的类 $S$,就称 T 刻画 $S$。例如,全序理论刻画全序结构的类,群论刻画群的类。常把论域是有穷集的结构称为有穷结构。

一阶逻辑有可靠且完全的公理系统,有许多良好的性质,是数理逻辑中发展得最为成熟的部分。在计算机科学和人工智能的各个领域,如数据结构理论、程序设计语言的形式语义,程序正确性证明、软件规范、程序变换、逻辑程序设计、定理的机器证明、常识表示和推理等方面,都广泛地应用了一阶逻辑的概念、理论和方法。

推理规则实质上是一种从其他语句产生新的谓词演算语句的机械方式。换句话来说,推理规则是基于给定逻辑断言的语法形式来产生新的语句。当推理规则从逻辑表达式集合 $S$ 产生的每一条语句都逻辑派生自 $S$ 时,我们就说这个推理规则是可靠的。

为了应用推理规则，如假言推理，推理系统必须能够判断两个表达式何时相同，也就是匹配。在命题演算中，这是显而易见的：两个表达式是匹配的，当且仅当它们在语句构成上相同。在谓词演算中，表达式中变量的∃使匹配两个表达式的过程变得复杂。全称量化允许用定义域中的项来替换全称量化变量。这需要一个决策处理来判断是否可以使变量替换产生的两个或更多个表达式相同(通常是为了应用推理规则)。

合一是一种判断什么样的替换可以使产生的两个谓词演算表达式匹配的算法。例如 $\forall X(\mathrm{man}(X)\rightarrow\mathrm{mortal}(X))$中的 $X$ 替换成了 man(socrates)中的 socrates。合一的另一个例子是在前面讨论哑元时看到的。因为 $p(X)$和 $p(Y)$是相同的，所以可以是用 $Y$ 替换 $X$ 使语句匹配。

对合一算法的要求是合一式要尽可能的通用，也就是要发现两个表达式的最一般合一式 mgu。如果 $s$ 是表达式 $E$ 的合一式，$g$ 是这个表达式集合的最一般表达式，那么对于应用到 $E$ 的 $s$，∃另一个合一式 $s'$ 使 $Es=Egs'$，其中 $Es$ 和 $Egs'$ 是应用到表达式 $E$ 的合一式的组合。

合一对于任何使用谓词演算作为表示的人工智能解决器都是很重要的。合一确定了两个(或更多)谓词演算表达式等价的条件。这允许我们通过逻辑表示来使用像归结这样的推理规则，归结是经常需要回溯已找到所有可能解释的一种过程。下面给出合一函数 unify 的算法：

**算法 14.1**　function unify($E_1$, $E_2$)

```
begin
    case
        E1 和 E2 都是常量或链表为空:                    % 递归停止
            if E1 = E2 then return {}
            else return FAIL
        E1 是一个变量:
            if E1 出现在 E2 中,then return FAIL
            else return {E2/E1};
        E2 是一个变量:
            if E2 出现在 E1 中,then return FAIL
            else return {E1/E2};
        无论 E1 或 E2 是空的 then return FAIL           % 链表是不同大小的
        otherwise:
            begin
                HE1: = E1 的第一个元素;
                HE2: = E2 的第一个元素;
                SUBS1: = unify(HE1, HE2);
                if SUBS1: = FAIL then return FAIL;
                TE1: = apply(SUBS1, E1 的其余部分);
                TE2: = apply(SUBS1, E2 的其余部分);
                SUBS2: = unify(TE1, TE2);
                if SUBS2: = FAIL then return FAIL;
                else return composition(SUBS1, SUBS2)
        end
    end                                                 % end case
end
```

Horn 逻辑是一阶逻辑的子部分。原子公式称为正文字，原子公式的否定称为负文字，

它们统称为文字。例如,$P(a)$,$P(f(x))$是正文字,$\neg Q(a,f(x))$是负文字。形式为 $L_1 \vee \cdots \vee L_m$ 的公式称为子句,其中 $L_1,\cdots,L_m$ 是文字。用□表示不包含任何文字的空子句。它是一个永假式,即总是解释为假命题的公式。至多包含一个正文字的子句称为 Horn 子句,是逻辑学家霍恩(A Horn)于1951年首先研究的。例如,$P(a)$,$\neg P(a) \vee \neg Q(a,f(x))$,$P(a) \vee \neg P(x)$都是 Horn 子句,而 $P(a) \vee P(f(a))$却不是 Horn 子句。

Horn 逻辑是由 Horn 子句组成的一阶逻辑的子部分。它是逻辑程序设计的理论基础。在逻辑程序设计中采用了一种特殊的子句记号,用 $A_1,\cdots,A_k \leftarrow B_1,\cdots,B_n$ 表示子句 $A_1 \vee \cdots \vee A_k \vee \neg B_1 \vee \cdots \vee \neg B_n$。Horn 子句共分以下四种形式:

① $A \leftarrow B_1,\cdots,B_n$

② $A \leftarrow$,

③ $\leftarrow B_1,\cdots,B_n$

④ 空子句□。

其中①称为过程,②称为事实,③称为目标。

Horn 子句可以表达知识库和对知识库的询问。Prolog 语言是以 Horn 逻辑为基础的高级程序设计语言。一个 Prolog 程序实际上就是一个由事实和过程组成的 Horn 子句集,询问就是一个目标。运行一个 Prolog 程序就是通过计算机判断由程序和目标组成的 Horn 子句集是否有模型。

## 14.3 动态描述逻辑

### 14.3.1 描述逻辑

描述逻辑是一种基于对象的知识表示的形式化,也叫概念表示语言或术语逻辑。它是一阶逻辑的一个可判定的子集,具有合适定义的语义,并且具有很强的表达能力。一个描述逻辑系统包含四个基本组成部分:表示概念和关系的构造集;TBox 包含断言;ABox 实例断言;TBox 和 ABox 上的推理机制。一个描述逻辑系统的表示能力和推理能力取决于对以上几个要素的选择以及不同的假设。

描述逻辑中有两个基本元素,即概念和关系。概念解释为一个领域的子集;关系则表示在领域中个体之间所具有的相互关系,是在领域集合上的一种二元关系。

在一定领域中,一个知识库 $K = \langle T, A \rangle$由两个部分组成:TBox $T$ 和 ABox $A$。其中 TBox 是一个关于包含断言的有限集合,也称为术语公理的集合。包含断言的一般形式为 $C \sqsubseteq D$,其中 $C$ 和 $D$ 都是概念。ABox 是实例断言的有限集合,形为 $C(a)$,其中 $C$ 是一个概念,$a$ 是一个个体的名字;或者形为 $P(a, b)$,其中 $P$ 为一个原始关系,$a$, $b$ 为两个个体的名字。

一般地,TBox 是描述领域结构的公理的集合,它具有两方面的作用,一是用来引入概念的名称,二是声明概念间的包含关系。引入概念名称的过程即可以表示为 $A \doteq C$ 或者 $A \sqsubseteq C$,其中 $A$ 即为引入的概念。概念间的包含关系的断言可以表示为 $C \sqsubseteq D$。对于概念定义和包含关系,有:$C \doteq D \Leftrightarrow C \sqsubseteq D$ 且 $C \sqsubseteq D$。

ABox 是实例断言的集合,用于指明个体的属性或者个体之间的关系。它有两种形式

的断言，一是指明个体与概念间的属于关系，二是指明两个个体之间所具有的关系。在 ABox 中，对于论域中任意个体对象 a 和概念 C，关于对象 a 是否为概念 C 中的元素的断言称为概念实例断言，简称概念断言。若 $a \in C$，则记为 $C(a)$；若 $a \notin C$，则记为 $\neg C(a)$。

对于两个对象 $a$，$b$ 和关系 $R$，如果 $a$ 和 $b$ 满足关系 $R$，则称 $aRb$ 为关系实例断言，表示为 $R(a, b)$。关系断言是用来指明两个对象之间所满足的基本关系或者对象的属性，构成二元关系。

一般地，描述逻辑依据提供的构造算子，在简单的概念和关系上构造出复杂的概念和关系。通常描述逻辑至少包含以下算子：交（$\sqcap$），并（$\sqcup$），非（$\neg$），存在量词（$\exists$）和全称量词（$\forall$）。这种最基本的描述逻辑称之为 ALC。在 ALC 的基础上再添加不同的构造算子，则构成不同表达能力的描述逻辑。例如若在 ALC 上添加数量约束算子“$\leqslant$”和“$\geqslant$”，则构成描述逻辑 ALCN，这里不做详细介绍。ALC 的语法和语义如表 14.1 所示。

ALC 语义将概念解释为一定领域的子集，关系是该领域上的二元关系。形式上，一个解释 $I=(\Delta^I, \cdot^I)$ 由解释的领域 $\Delta^I$ 和解释函数 $\cdot^I$ 所构成，其中解释函数把每个原子概念 $A$ 映射到 $\Delta^I$ 的子集，而把每个原子关系 $P$ 映射到 $\Delta^I \times \Delta^I$ 的子集。

(1) 一个解释 $I$ 是包含断言 $C \sqsubseteq D$ 的模型，当且仅当 $C^I \subseteq D^I$。

(2) 解释 $I$ 是 $C(a)$ 的模型，当且仅当 $a \in C^I$；$I$ 是 $P(a, b)$ 的模型，当且仅当 $(a, b) \in P^I$。

(3) 解释 $I$ 是知识库 $\mathcal{K}$ 的模型，当且仅当 $I$ 是 $\mathcal{K}$ 中每个包含断言和实例断言的模型。

(4) 若 $\mathcal{K}$ 有模型，则称 $\mathcal{K}$ 是可满足的。

(5) 若断言 $\delta$ 对于 $\mathcal{K}$ 的每个模型是满足的，则称 $\mathcal{K}$ 逻辑蕴含 $\delta$，记为 $\mathcal{K} \models \delta$。

(6) 对概念 $C$，若 $\mathcal{K}$ 有一个模型 $I$ 使得 $C^I \neq \varnothing$，则称 $C$ 是可满足的。知识库 $\mathcal{K}$ 中的概念 $C$ 的可满足性可以逻辑表示为 $\mathcal{K} \not\models C \sqsubseteq \bot$。

**表 14.1　ALC 的语法和语义**

| 构造算子 | 语　法 | 语　义 | 例　子 |
|---|---|---|---|
| 原子概念 | $A$ | $A^I \subseteq \Delta^I$ | Human |
| 原子关系 | $P$ | $P^I \subseteq \Delta^I \times \Delta^I$ | has-child |
| 顶部 | $\top$ | $\Delta^I$ | True |
| 底部 | $\bot$ | $\Phi$ | False |
| 交 | $C \sqcap D$ | $C^I \cap D^I$ | Human $\sqcap$ Male |
| 并 | $C \sqcup D$ | $C^I \cup D^I$ | Doctor $\sqcup$ Lawyer |
| 非 | $\neg C$ | $\Delta^I - C^I$ | $\neg$ Male |
| 存在量词 | $\exists R.C$ | $\{x \mid \exists y, (x, y) \in R^I \wedge y \in C^I\}$ | $\exists$ has-child. Male |
| 全称量词 | $\forall R.C$ | $\{x \mid \forall y, (x, y) \in R^I \Rightarrow y \in C^I\}$ | $\forall$ has-child. Male |

关于描述逻辑中的基本推理问题，主要包括概念的可满足性、概念的包含关系、实例检测、一致性检测等，其中概念的可满足性问题是最基本的问题，其他的推理基本上都可以转化为概念的可满足性问题。

在描述逻辑中，可以利用下述性质对推理问题进行约简，转化为概念的可满足性问题，进而将推理问题进行简化。对于概念 $C$，$D$，有如下命题成立。

(1) $C \sqsubseteq D \Leftrightarrow C \sqcap \neg D$ 是不可满足的。

(2) $C \doteq D$ 是等价的 $\Leftrightarrow (C \sqcap \neg D)$ 与 $(D \sqcap \neg C)$ 都是不可满足的。

(3) $C$ 与 $D$ 是不相交的⇔$C \sqcap D$ 是不可满足的。

## 14.3.2 动态描述逻辑 DDL

由于动态描述逻辑是在传统描述逻辑的基础上扩充得到的,而传统描述逻辑有很多种类,本节以最小的描述逻辑 ALC 为基础来研究动态描述逻辑 DDL。

### 1. DDL 的语法

**定义 14.1** 在 DDL 的语言中包括以下基本符号:

- 概念名:$C_1$, $C_2$, …
- 关系名:$R_1$, $R_2$, …
- 个体常元:$a$, $b$, $c$, …
- 个体变元:$x$, $y$, $z$, …
- 概念运算:¬, ⊓, ⊔ 以及量词:∃, ∀
- 公式运算:¬, ∧, → 以及量词 ∀
- 动作名:$A_1$, $A_2$, …
- 动作构造:如;(合成),∪(交替),*(反复),?(测试)
- 动作变元:$\alpha$, $\beta$, …
- 公式变元:$\varphi$, $\psi$, $\pi$, …
- 状态变元:$u$, $v$, $w$, …

**定义 14.2** 在 DDL 中,概念定义如下:

(1) 原子概念 $P$、全概念⊤、和空概念⊥都是概念。

(2) 如果 $C$ 和 $D$ 是概念,则$\neg C$,$C \sqcap D$,$C \sqcup D$ 都是概念。

(3) 如果 $R$ 为关系,$C$ 为概念,则 $\exists R.C$,$\forall R.C$ 都是概念。

(4) 如果 $C$ 是概念,$\alpha$ 是动作,则$[\alpha]C$ 也是概念。

**定义 14.3** DDL 的公式定义如下,其中 $C$ 为任意概念,$R$ 为关系,$a$, $b$ 为个体常元,$x$,$y$ 为个体变元,$\alpha$ 是动作。

(1) 形如 $C(a)$,$R(a, b)$和$[\alpha]C(a)$的表达式称为断言公式,它们是不带变元的。

(2) 形如 $C(x)$,$R(x, y)$和$[\alpha]C(x)$的表达式称为一般公式,它们是带变元的。

(3) 断言公式和一般公式都是公式。

(4) 如果 $\varphi$ 和 $\psi$ 是公式,则$\neg\varphi$,$\varphi \wedge \psi$,$\varphi \rightarrow \psi$,$\forall x\varphi$ 都是公式。

(5) 如果 $\varphi$ 是公式,则$[\alpha]\varphi$ 也是公式。

**定义 14.4** 形如$\{a_1/x_1, \cdots, a_n/x_n\}$的有穷集合称为一个实例代换,其中 $a_1$, …, $a_n$ 为个体常元,称为代换项,$x_1$, …, $x_n$ 为个体变元,称为代换基,它们满足 $x_i \neq x_j$,$i$, $j \in \{1, \cdots, n\}$。

**定义 14.5** 设 $\varphi$ 为一公式,$x_1$, …, $x_n$ 为出现在 $\varphi$ 中的个体变元,$a_1$, …, $a_n$ 为个体常元,令 $\varphi'$为 $\varphi$ 通过实例代换$\{a_1/x_1, \cdots, a_n/x_n\}$而得到的公式,则称 $\varphi'$为公式 $\varphi$ 的实例公式。

**定义 14.6** DDL 中条件(Condition)定义如下,其中 $N_C$ 表示个体常元的集合,$N_X$ 表示个体变元的集合,$N_I$ 是 $N_C$ 和 $N_X$ 的并,即 $N_I = N_C \cup N_X$:

$\forall C$,$C(p)$,$R(p,q)$,$p=q$,$p \neq q$。其中 $p,q \in N_I$,$C$ 是 DDL 的概念,$R$ 是 DDL 的关系。

**定义 14.7**　一个动作描述是一个形如 $A(x_1,\cdots,x_n)\equiv(P_A,E_A)$的表达形式，其中：

(1) $A$ 为动作名：指示动作表示符。

(2) $x_1$，…，$x_n$ 为个体变元，指定动作的操作对象，因此也称为操作变元。

(3) $P_A$ 为前提公式集(Pre-conditions)，指定动作执行前必须满足的前提条件，即 $P_A=\{con \mid con\in condition\}$。

(4) $E_A$ 为结果公式集(Post-conditions)，指定动作执行后得到的结果集，$E_A$ 是序对 $head/body$ 的集合，其中 $head=\{con \mid con\in condition\}$，$body$ 是一个条件。

**说明：**

(1) 动作定义了状态间的转换关系，即一个动作 $A$ 将一个状态 $u$ 转换成状态 $v$，如果在状态 $u$ 下应用动作 $A$ 则产生状态 $v$。这种转换关系依赖于状态 $u$，$v$ 是否分别满足动作 $A$ 的前提公式集(*pre-conditions*)和结果公式集(*post-conditions*)，记作 $uT_Av$。

(2) 因为动作 $A$ 发生以前的状态也可以影响动作 $A$ 的结果，因而前提公式与结果公式在描述上有些不同。对于结果公式 $head/body$，如果 $head$ 中的每个条件在状态 $u$ 中满足，则 $body$ 中的每个条件在状态 $v$ 中满足。例如，卖自行车的动作(Bicycle-selling)可以描述如下：$(\{owns(a_1,b), wants(a_2,b), owns(a_2,p), bicycle(b)\}, \{\Phi/owns(a_2,b), \{bad(b)\}/\neg happy(a_2), \Phi/owns(a_1,p), \{bad(b)\}/happy(a_1)\})$。

**定义 14.8**　设 $A(x_1,\cdots,x_n)\equiv(P_A,E_A)$为一个动作描述，$A(a_1,\cdots,a_n)$是在 $A(x_1,\cdots,x_n)$上经过实例代换$\{a_1/x_1,\cdots,a_n/x_n\}$而得到的，则称 $A(a_1,\cdots,a_n)$为 $A(x_1,\cdots,x_n)$的动作实例，并称 $A(a_1,\cdots,a_n)$为原子动作，$P_A(a_1,\cdots,a_n)$称为动作 $A(a_1,\cdots,a_n)$的前提集，$E_A(a_1,\cdots,a_n)$称为动作 $A(a_1,\cdots,a_n)$的结果集。

**定义 14.9**　DDL 的动作定义如下：

(1) 原子动作 $A(a_1,\cdots,a_n)$是动作。

(2) 如果 $\alpha$ 和 $\beta$ 为动作，则 $\alpha;\beta$，$\alpha\cup\beta$，$\alpha*$ 都是动作。

如果 $\varphi$ 为断言公式，则 $\varphi?$ 也是动作。

有关动态描述逻辑 DDL 的语义请参阅文献[933]。

## 14.4　归纳逻辑

一般来说，演绎推理的前提蕴涵结论，前提为真，结论就一定为真。归纳推理的前提与结论之间不具有这种蕴涵关系，归纳推理的结论超出了前提的范围，因而当前提为真时，结论不一定为真。归纳逻辑是关于或然性推理的逻辑。

归纳逻辑最早的思想萌芽可以追溯到古代。在古代中国、古代印度、古希腊的逻辑研究中都曾经涉及归纳。就西方逻辑传统而言，最早接触归纳的是德谟克利特。据记载，德谟克利特有一部叫做《规范》的著作，其中研究了归纳、类比等问题。可惜这部著作已失传。德谟克利特之后，苏格拉底(Soocrates)、柏拉图(Plato)都使用了归纳的认识方法。亚里士多德对归纳做了较为详细的研究。他研究了完全归纳推理，简单枚举归纳推理和作为科学认识方法的归纳法。

16 世纪末到 18 世纪初，古典归纳逻辑应运而生。古典归纳逻辑的奠基人是英国哲学家弗兰西斯·培根。培根主要的哲学逻辑著作是《新工具》。培根把归纳法分别介绍。第一

步,尽量全面地搜集经验材料,归纳法要以丰富的客观材料为基础,观察和实验则是收集经验材料的方法,因而也是归纳法的基础。第二步,对搜集来的材料进行整理、排列。这里,培根提出了著名的“三表法”。三表法是整理、分析、比较材料的方法,也是寻求因果联系的方法。第一表是“存在和具有表”,给定存在的一个性质,首先要把所有已知的,虽然在材料上很不相同,但在这个性质上是一致的例证收集起来,排列出来。也就是从某些不同的事物中找出它们的共同点,以便发现所考察性质的原因。第二表是“缺乏表”。列出给定性质不存在的各种例证。当给定性质不存在时,它的原因也应当不存在。第三表是“程度表”,或“比较表”。列出给定性质出现的程度不同的各种例证,找出随性质的增减而增减的现象,这才是确立该性质的真正原因。培根用“三表法”来整理经验材料,判明因果联系,但他并不认为三表法就是归纳法,而是把它作为归纳推理的准备,作为归纳推理的一个环节。第三步,进行真正的归纳。

19 世纪的英国逻辑学家穆勒是古典归纳逻辑的集大成者。他继承并发展了培根、赫舍尔、惠威尔的归纳学说,在其名著《逻辑体系》中论述了确定现象间因果联系的五种方法,即归纳五法。它们是契合法、差异法、契合差异并用法、剩余法和共变法。这五种方法是培根“三表法”和赫舍尔因果决定法的精密化与规范化。

正当古典归纳逻辑向前发展的时候,18 世纪的英国哲学家休谟对归纳推理的合理性提出了质疑。休谟提出的问题是:归纳法具有理性的依据吗?如何为归纳法的合理性进行辩护?休谟本人的回答是:为归纳法的合理性进行辩护是不可能的,因此归纳法没有合理性,只不过是人的一种心理本能。休谟的理由大致是,一切推理可以分为两类:一类是关于观念间的推理,具有必然性;另一类是关于经验事实的推理,具有或然性。归纳法是要根据过去发生的事情推断将来要发生的事情,既然过去和将来之间没有逻辑上的必然性,所以不能用前一种推理为它进行辩护,但也不能用后一种推理为它进行辩护,否则就会出现循环论证。

除了休谟问题外,现代归纳逻辑还面临若干悖论,其中包括认证悖论(乌雅悖论)、绿蓝悖论(新归纳之谜)和抽彩悖论,它们分别由当代逻辑学家和哲学家亨佩尔(C G Hempel)、古德曼(N Goodman)和凯伯格(H E Kyburg)提出。这些悖论的共同特点是,从人们通常公认的原则或原理出发,却得出逻辑矛盾或与常识相违的结论。对于这些悖论能否给出恰当的解决,是衡量一种归纳理论是否恰当的重要标志。

1921 年,英国著名的经济学家凯恩斯(J M Keynes)将概率理论与归纳逻辑相结合,建立了第一个概率逻辑系统,这标志着现代归纳逻辑的产生。此后,逻辑学家们纷纷提出各自的归纳逻辑系统。现代归纳逻辑的特点是,第一,把概率概念引进归纳逻辑,人们充分认识到归纳推理的或然性,试图从量上刻画这种或然性,很自然地采用了概率概念,现代归纳逻辑的研究几乎都是结合概率、统计理论进行的;第二,不再把归纳看作发现和证明普遍性命题(规律、定律)的活动,而把它看作检验假说的活动,归纳法的职能不是发现全称命题更不能证明它,归纳法主要是通过检验来决定一个假说是否可被接受;第三,数理逻辑即现代演绎逻辑的方法对归纳逻辑的研究产生很大影响,公理化形式化的方法被引入归纳逻辑的研究,产生了许多不同类型的形式化的归纳逻辑系统。

20 世纪 60 至 70 年代出现了一种新的思潮即局部归纳逻辑。局部归纳逻辑不同于整体归纳逻辑的地方在于,它不要求对一切非演绎的原则或知识进行辩护,而只要求对那些在

科学家们看来已经成为问题的原则或知识进行辩护。尽管局部归纳逻辑对于现代归纳逻辑的发展起了相当大的促进作用，但是如此宽泛的局部化使其哲学价值受到怀疑。主观主义亦即贝叶斯主观概率归纳逻辑走了一条介于局部归纳逻辑和整体归纳逻辑之间的道路，贝叶斯主观概率归纳逻辑代表着现代归纳逻辑的发展趋势。下面就对有关问题分别加以简要的讨论。

## 14.4.1　经验主义概率归纳逻辑

经验主义概率归纳逻辑主要是由莱欣巴赫（H Reichenbach）于 20 世纪 30 年代提出的，后由萨尔蒙（W Salmon）等人给以进一步的发展。在此理论中，概率被定义为相对频率的极限。莱欣巴赫在 1935 年发表的《概率理论》一书中构造了一个概率逻辑系统，他的目的是要论证归纳推理的合理性，解决休谟问题。莱欣巴哈首先建立了一个概率演算的公理系统，这个系统是以狭谓词演算为基础，增加新的符号和公式构造出来的。他引进一个新的概念——概率蕴涵，用符号 $\ni_p$ 表示之。$p$ 是一个变元，取[0，1]区间的任何有理数为值。利用这个符号可以构成表达概率语句的公式：

$$(i)(x_i \in A \ni_p y_i \in B) \tag{14.1}$$

其中$(i)$是全称量词，$A$、$B$ 表示类，概率蕴涵可以看作类之间的一种关系。他把类 $A$ 称为参考类，把类 $B$ 称为属性类。让类 $A$ 和类 $B$ 的元素都排成序列，并约定两个序列的元素之间有一一对应关系。上述概率语句可以简写为

$$P(A,B) = p \tag{14.2}$$

莱欣巴赫的概率演算有五条关于概率的公理，在狭谓词演算中加入这五条公理，可以推演出数学概率论中的全部结果。

接着莱欣巴赫对概率作了频率解释。相对频率用 $F^n(A,B)$表示。

$$F^n(A,B) = \frac{N^n(A \cdot B)}{N^n(A)} \tag{14.3}$$

其中

$$N^n(A) =_{df} N_{i=1}^n(x_i \in A)$$
$$N^n(A \cdot B) =_{df} N_{i=1}^n(x_i \in A) \cdot (y_i \in B)$$

它们分别定义为序列 $x_1, x_2, \cdots, x_n$ 中属于参考类 $A$ 的数目，以及由序列 $x_1, x_2, \cdots, x_n$ 与序列 $y_1, y_2, \cdots, y_n$ 的对应元素组成的 $x_1, x_2, ..., x_n$ 序对 $x_1y_1, x_2y_2, \cdots, x_ny_n$ 中满足 $x_i \in A$ 且 $y_i \in B$ 的序对数目。就序对 $x_iy_i$ 而言，当 $n \to \infty$，相对频率 $F^n(A,B)$趋于极限 $P$，就称该序对中，由 $A$ 到 $B$ 的概率，表示为

$$P(A,B) = \lim_{n \to \infty} F^n(A,B) \tag{14.4}$$

莱欣巴赫用概率的频率解释来处理归纳推理。实际上，概率演算和概率逻辑都是演绎性质的，唯有求初始概率的过程本质上是归纳的过程。如何求初始概率？在频率解释下，概率是相对频率的极限，亦即极限频率，莱欣巴哈主张通过考察相对频率来认定（Posit）极限频率，即通过考察相对频率 $F^1(A,B), F^2(A,B), \cdots, F^n(A,B)$，认定极限频率随 $n \to \infty$将趋于其中的某个 $F^i(A,B)$，$1 \leqslant i \leqslant n$。这样做出的认定有可能不正确，这时需要继续考察，重新认定某个 $F^i(A,B)$是极限频率。莱欣巴赫把简单枚举法的合理性描述为，如果极限频率

$\lim_{n\to\infty} F^n(A)$存在,那么坚持运用简单枚举法,不断修改原有结论,作出新的认定,就能由相对频率求出极限频率。但是极限频率是否存在?莱欣巴赫认为,我们不能说它存在,也不能说它不存在。如果极限频率存在,坚持运用归纳推理,就一定能找到它。如果极限频率不存在,运用归纳法会出错,但这时别的方法也会出错。与别的方法相比,归纳法是最简单最好的方法。这就是莱欣巴赫为解决休谟问题而提出的方案。

### 14.4.2 概率逻辑理论

逻辑贝叶斯派中最有影响的是卡尔纳普(R Carnap)的概率逻辑理论。卡尔纳普把概率区分为概率$_1$ 和概率$_2$。概率$_2$ 是频率解释,概率$_1$ 是逻辑概率,表示证据对假说的确证程度。卡尔纳普认为,逻辑概率概念是一切归纳推理的基础,如果能找到一种令人满意的逻辑概率的定义和理论,那么就可以为最终解决归纳推理的争论提供一个清楚而合理的基础。卡尔纳普以严密的逻辑形式,构造了一个概率逻辑系统。他假定一个语言 $L$ 的初始命题函项具有形式 $P_i(x_i)$,把 $L$ 的所有初始谓词和个体常元分别代入这个函数项,就可以形成 $L$ 的所有基本句子。再构造这样的合取式,使得 $L$ 中的每一个基本句子或其否定在其中出现,这样的合取式称为状态描述。任何两个状态描述都是互不相容的。如果以某种方式将非负的数值赋予每一个状态描述,并且使得这些数值的总和为 1,那么就可以得到状态描述的量度。一个命题 $P$ 的值域定义为使命题 $P$ 成立的所有那些状态描述的集合。命题 $P$ 可以表示为所有这些状态描述的析取,该命题的量度等于这些状态描述的量度之和。用 $m(e)$表示命题 $e$ 的量度,称 $m(e)$为 $e$ 的先验概率。用 $c(h, e)$表示命题 $e$ 对命题 $h$ 的确证度,就有

$$c(h,e) = \frac{m(h\Lambda e)}{m(e)} \tag{14.5}$$

也就是说 $h$ 在 $e$ 基础上的确证度定义为 $h$ 与 $e$ 的合取的先验概率除以 $e$ 的先验概率所得的商。称 $c(h, e)$为确证函数或 $c$ 函数。现在,问题的关键是如何给状态描述指派量度。一种做法是,根据无差别原则,给 $L$ 语言中的每一个状态描述以同等的量度。但如果这样做,那么将永远是 $c(h, e)=m(h)$,即 $h$ 在证据 $e$ 基础上的确证度永远等于它的先验概率。这显然是卡尔纳普不希望得到的结果,于是他引进“结构描述”这一概念。所谓结构描述是通过交换个体常元的名称就能够互相转化的所有不同的状态描述的析取。运用无差别原则,先赋予所有的结构描述以同样的量度,然后再次运用无差别原则,把这一度量平分给这一结构描述中所包含的状态描述。这种做法避免了前两种做法的困难。但这种方法也远不是理想的。起初卡尔纳普相信只有一种 $c$ 函数是合理的,后来他发现,若引进参数 $\lambda$,$\lambda$ 取 0 到$+\infty$的任意值,则每一个 $\lambda$ 值确定一种 $c$ 函数,根据 $\lambda$ 的不同取值,可以得到确证函数的连续集合。

### 14.4.3 主观贝叶斯概率

对概率的频率解释与逻辑解释都是对概率的客观量度。事实上,对随机事件出现的预测离不开主观的信念与期望,离不开人的心理因素的作用。因此还存在着概率的主观解释,又称为主观(或个人)概率,这是对信念与期望的度量。早在 1926 年,拉姆齐(F P Ramsey)已对与归纳推理有关的主观概率进行过研究。这对后来这方面的研究是有影响的。卡尔纳

普虽然力图从客观方面研究概率归纳逻辑，但仍无法完全避免主观概率问题。1962 年，他将归纳逻辑与合理决策联系起来研究了主观概率问题。这种主观概率是一种与效用有关的置信度。某 $X$ 在 $T$ 时刻对语句 $H$ 的置信度称为置信函数，以 $\mathrm{Crx}\ (h,T)$表示。如果 $x$ 在 $T$ 时刻的全部观察知识为 $A$，且 $X$ 对 $H$ 的置信度只与 $A$ 有关，则称为可信函数，以 $\mathrm{Credx}(H/A)$表示。他找出 $\mathrm{Crx}(h,T)$与 $\mathrm{Credx}(H/A)$之间的一定关系，即

$$\mathrm{Crx}(h,T) = \mathrm{Credx}(h,K(T))$$

其中 $K(T)$表示 $T$ 以前全部知识的合取。他还将逻辑概率解释为合理的可信函数，试图以此将归纳逻辑中的客观因素与主观因素协调起来，对确认度 $C$ 亦即逻辑概率作出了另一种解释，将概率逻辑的研究推进了一步。

主观贝叶斯主义者认为，对同一事件或命题，主体可以自由地选择自己的置信度，也就是说，可以有 0 到 1 之间的任何置信度。主体对相关命题的置信度集合应当是一致的，当且仅当一个人的置信度集合满足概率论的公理和规则时，他的置信度集合是一致的。例如，对 $P$ 的置信度加上对非 $P$ 的置信度等于 1 等。如果某个人的置信度集合不一致，那么在打赌时，狡猾的打赌者会专门找他打赌，使得他在任何情况下都输。主体的置信度集合应当满足概率计算的全部规则，这一要求对打赌不会总是失败是充分而且必要的，莱维(G Levi)把这一判断称为“荷兰赌定理”(Dutch Book theorem)。实际上，并不是所有人的置信度集合都是一致的，因此，主观贝叶斯派讨论的是合乎理性的，理想的人的置信度，称为“合理置信度”。

对同一个命题或事件，合理的置信度集合不是唯一的，如果两个置信度集合都满足概率论基本定律所表达的数量关系，就不能说它们中一个比另一个更合理。这个结果显然是不能令人满意的。因此，菲尼蒂引进了“可换事件的序列”这一概念，提出所谓“可换事件”的定理。根据菲尼蒂的理论，如果主体同意某个序列的事件是可换的，并且他们愿意从经验中学习，那么，尽管开始的时候他们对同一事件的出现具有不同的置信度，但在给定这个序列的一个足够长的部分做证据的情况下，他们对未来事件出现的概率将会有非常一致的看法。

### 14.4.4 条件化归纳逻辑

条件化归纳逻辑是英国哲学家布罗德(C D Broad)首先提出的，后来冯·赖特(G H von Wright)做了更加系统的研究和发展。所谓条件化归纳逻辑是指对条件蕴涵式的前、后件以及前、后件之间的关系进行严密的逻辑分析，使之构成一个较严格的逻辑系统，然后以此为工具，对穆勒的契合法、差异法、契合差异并用法加以条件化的阐释和处理，使之成为一种以演绎条件句为基础的归纳逻辑理论。穆勒的归纳方法是寻求现象间因果联系的方法，而在日常语言中，“原因”一词有许多不同的含义，穆勒的归纳方法也缺乏逻辑的精确性和形式的严密性。赖特等人用一般的条件蕴涵式来表达因果关系，定义了充分条件、必要条件、充分必要条件等概念，提出并证明了与条件关系以及与这些关系的推导有关的一系列的定理，构成条件逻辑系统。然后用条件逻辑的基本原理来处理穆勒的排除归纳法。他们把不同情况的原因化归为充分条件、必要条件、充要条件三种，把一些因果关系的陈述用全称条件语句的形式表达出来。这样，排除归纳法就是确定某个事物存在的条件。给定一个被研究的现象，列出该现象存在的可能的充分条件(或必要条件，或充要条件)的一览表，用适当的方法排除表中除一个条件之外的其他可能的条件，最后剩下的条件就是被研究现象的充

分条件(或必要条件,或充要条件)。排除可能条件的过程是演绎推理的过程。条件化归纳逻辑使穆勒的排除归纳法有了较为严密的逻辑形式和结构,但用形式为$(x)(Fx \supset Gx)$这类全称命题来表达自然规律或因果律是否恰当却是一个值得讨论的问题,许多逻辑学家建议引进新的联结词来表达因果陈述。另外,使用排除归纳法首先要列出可能的条件的一览表,并假设被寻求的那个条件也在其中。如何列出这样的表,也是一个值得讨论的问题。赖特的处理方法既烦锁又复杂,实际上,这是一个经验的问题,应当由具体的研究者决定。再有,即使研究者根据自己的经验列出了这样的表,并使用排除法得到了最后的结果,他也不能绝对确信这个结论是正确的。

### 14.4.5 非帕斯卡概率归纳逻辑

非帕斯卡概率论这个概念首先由科恩于 1977 年正式提出,但对它的研究可以追溯到沙克尔(G Shackle)[955]。所谓帕斯卡(Pascal)概率论就是经典概率论;它有一条定理,即$P(\neg H)=1-P(H)$,此定理称为“否定律”,也称为“互补律”。但是,此定理在非帕斯卡概率论中不成立。

沙克尔在《经济学中的期望值》一书中,首次提出非分布不确定性概念,并对它作出主观主义的解释,在此基础上建立了非帕斯卡概率演算规则。它的一般表述是:令 $h$ 为任一事件或假说,$M>0$,$P_I$ 是非帕斯卡概率函数,它满足下列演算规则:

(1) $M \geqslant P_I(h) \geqslant 0$。

(2) 如果 $P_I(h)>0$,那么 $P_I(\neg A)=0$(否定律)。

(3) $P_I(h_1 \wedge h_2)=\min\{P_I(h_1), P_I(h_2)\}$(合取律)。

(4) $P_I(h_1 \vee h_2)=\max\{P_I(h_1), P_I(h_2)\}$(析取律)。

将上述规则同帕斯卡概率演算规则相比,最为明显的区别在于,$P_I$ 函数不满足帕斯卡概率函数定义中的可加性条件,也不满足帕斯卡概率函数的互补律。值得一提的是,规则(2)使非帕斯卡概率演算规则与模糊推理规则绝然不同。

科恩的非帕斯卡概率归纳逻辑是对其归纳支持理论的简单扩展,即把一个普遍概括的归纳支持度移植到它的某个特殊事例上。归纳支持理论是以相关变项法为其语义模型的,因此,科恩的非帕斯卡概率同支持度一样也是分级的,而非连续的。

以 $x$ 、$y$、$z$ 等表示个体变元,$H$、$I$、$P$ 等表示命题变元,$R$、$S$ 等表示一阶谓词符号,联结词与量词的符号和解释与命题逻辑相同。另引入强度不同的模态算子以刻划归纳支持与非帕斯卡型归纳概率的等级。以□表示模态常项,模态分级如下:

$\Box^d H$ 表示 $H$ 为逻辑真(逻辑必然性)

以下为物理必然性,分级为:

$\Box^e H$ 表示 $H$ 得到充分支持,归纳概率值最高

$\Box^i H$ 表示 $H$ 得到第 $i$ 级支持,归纳概率值为 $i$

$\Box^o H$ 表示 $H$ 未得到支持,归纳概率值为 $o$

其中,$e$、$I$、$o$ 分别表示通过检验的相关变量数。

科恩曾指出了帕斯卡型的数学概率不宜运用于归纳概率的描述,而采取非帕斯卡型概率。非帕斯卡型概率定义为:定义在某个集合$\{A\}$上的函数 $P_I$ 称为非帕斯卡型概率,如果对一切 $A\in\{A\}$,它满足以下条件:

(1) $0 \leqslant P_I(A) \leqslant 1$。

(2) 若 $P_I(A) > 0$,则 $P_I(\neg A) = 0$。

科恩的归纳支持与归纳概率要求满足:

(1) 合取原理,若 $H$ 与 $H'$ 的相关变量至少部分重合,且 $S[H',E] \geqslant S[H,E]$,则 $S[H \wedge H',E] = S[H,E]$。相应地,若 $P_i(S', R) \geqslant P_i(S, R)$,则 $P_I(S' \wedge S, R) = P_I(S, R)$。

(2) 否定原理,只要归纳概括 $\forall x(Rx \to Sx)$ 在实验检验中为非零支持度,从而 $P_I(S, R) > 0$,则 $P_I(\neg S, R) = 0$。即互补关系在此不成立。

由于非帕斯卡概率不满足经典概率的互补律,这使得任何一个假设如果曾经获得大于 0 的支持度,那么它就永远不会被彻底否定。更有甚者,如果一个假设曾经在实验 $t_i$ 中获得较高的支持度如 4/5,那么,$t_i$ 以后的任何否证性实验都不能使之降低支持度。这一结论是与科学检验的实际情况相违背的。总之,与帕斯卡概率论相比,非帕斯卡概率论以及相应的归纳逻辑无论从语法上还是从语义上都显得不够成熟,亟待改进和发展。

## 14.5 范畴论

范畴论(Category Theory)是抽象地处理数学结构以及结构之间联系的一门数学理论,以抽象的方法来处理数学概念,将这些概念形式化成一组组对象及态射。1945 年,艾伦伯格(S Eilenberg)和麦克兰恩(S MacLane)引入范畴、函子和自然变换。这些概念最初出现在拓扑学,尤其是代数拓扑学里,在同态(具有几何直观)转化成同调论(公理化方法)的过程中起了重要作用。范畴自身亦为一种数学结构。函子(Functor)将一个范畴的每个对象(Object)和另一个范畴的对象相关连起来,并将第一个范畴的每个态射(Morphism)和第二个范畴的态射相关连起来。一个范畴 $C$ 包含两个部分:对象和态射。

态射是两个数学结构之间保持结构的一种过程抽象。在集合论中,态射就是函数;在群论中,它们是群同态;而在拓扑学中,它们是连续函数;在泛代数(Universal Algebra)的范围,态射通常就是同态。

常见的态射类型如下。

(1) 同构(Isomorphism):令 $f: X \to Y$ 为一个态射,若存在态射 $g: Y \to X$ 使得 $f \circ g = \mathrm{id}_Y$ 和 $g \circ f = \mathrm{id}_X$ 成立,则 $f$ 称为一个同构。$g$ 称为 $f$ 的逆态射,逆态射 $g$ 如果存在就是唯一的,而且显而易见 $g$ 也是一个同构,其逆为 $f$。两个对象之间有一个同构,那么这两个对象称为同构的或者等价的。同构是范畴论中态射的最重要的种类。

(2) 满同态(Epimorphism):一个态射 $f: X \to Y$ 称为一个满同态,如果对于所有 $Y \to Z$ 的态射 $g_1, g_2 g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$ 成立。这也称为 epi 或 epic。具体范畴中的满同态通常是满射(Surjective)函数,虽然并不总是这样。

(3) 单同态(Monomorphism):态射 $f: X \to Y$ 称为单同态,如果对于所有 $Z \to X$ 的态射 $g_1, g_2, f \circ g_1 = f \circ g_2 \Rightarrow g_1 = g_2$ 成立。它也称为 mono 或者 monic。具体范畴中的单同态通常为单射(Injective)函数。

(4) 双同态(Bimorphism):若 $f$ 既是满同态也是单同态,则称 $f$ 为双同态(Bimorphism)。

如果在一个范畴中每个双同态都是同构,则这个范畴称为一个平衡范畴。例如,集合是

一个平衡范畴。

(5) 自同态(Endomorphism):任何态射 $f$: $X \to X$ 称为 $X$ 上的一个自同态。

(6) 自同构(Automorphism):若一个自同态也是同构的,那么称为自同构。

(7) 若 $f$: $X \to Y$ 和 $g$: $Y \to X$ 满足 $f \circ g = \mathrm{id}_Y$ 可以证明 $f$ 是满的而 $g$ 是单的,而且 $g \circ f$: $X \to X$ 是幂等的。这种情况下,$f$ 和 $g$ 称为分割(Split),$f$ 称为 $g$ 的收缩(Retraction),而 $g$ 称为 $f$ 的截面。任何既是满同态又是分割单同态的态射,或者既是单同态又是分割满同态的态射必须是同构。

每个态射 $f$: $A \to B$,其中 $A$,$B$ 是 $C$ 中的对象。$A$ 是 $f$ 的定义域,记为 dom $f$;$B$ 是 $f$ 的值域,记为 cod $f$。如果两个态射 $f$,$g$ 适合 dom $g$ = cod $f$,则可以结合 $g$ 和 $f$ 得到 $g \square f$: $A \to B'$,这里,$A$ = dom $f$,$B'$ = cod $g$。有时把 $g \square f'$ 记为 $gf$。

**定义 14.10** 范畴 $C$

(1) 一族对象 obC。

(2) 任意一对对象 $A$, $B$,对应一个集合 $C(A,B)$,其元素称为态射使得当 $A \neq A'$ 或者 $B \neq B'$ 时,$C(A, B)$ 与 $C(A', B')$ 不交。

组成,满足下面条件:

(a) 复合律:若 $A,B,C \in \mathrm{ob}C$,$f \in C(A, B)$,$g \in C(B, C)$,则存在唯一的 $gf \in C(A, C)$,称为 $f$ 与 $g$ 的复合。

(b) 结合律:若 $A,B,C,D \in \mathrm{ob}C$,$f \in C(A, B)$,$g \in C(B, C)$,$h \in C(C,D)$,则有 $h(gf) = (hg)f$。

(c) 单位态射:每一个对象 $A$,存在一个态射 $1_A \in C(A,A)$,使得对任意的 $f \in C(A, B)$ 及 $g \in C(C, A)$ 有

$$f1_A = f, \quad 1_{Ag} = g$$

关于范畴的定义在一些文献中有着不同的表达形式,一些文献中的范畴定义不要求任意两个对象之间的态射的全体是一个集合。在范畴论中记号约定如下:用花体字母如 $\mathscr{D}$,$\mathscr{C}$ 等表示范畴,范畴中的对象用大写英文字母表示,而态射用小写英文字母或小写希腊字母表示。设 $\mathscr{C}$ 是一个范畴,$\mathscr{C}$ 的态射的全体记作 Mor$\mathscr{C}$。

下面列出一些范畴例子,这里只给出对象和态射。

- 集合范畴 Set(在某个给定的集合论模型中),其对象为集合,态射为映射。
- 群范畴 Gp,其对象为群,态射为群同态。类似地有 Abel 群范畴 AbGp,环范畴 Rng 和 R 模范畴 $\mathrm{Mod}_R$。
- 拓扑空间范畴 Top,其对象为拓扑空间,态射为连续映射。类似地有拓扑群范畴 TopGp,其对象为拓扑群,态射为连续的群同态。可微流形为对象光滑,映射为态射的范畴 Diff。
- 拓扑空间同伦范畴 Htop,其对象为拓扑空间,态射为连续映射的同伦等价类。
- 点拓扑空间范畴 Top *,其对象为序对 $(X,x)$,其中 $X$ 是非空拓扑空间,$x \in X$,态射为保点连续映射($f$: $(X,x) \to (Y,y)$ 称为保点连续映射当且仅当 $f$: $X \to Y$ 是连续映射并且满足 $f(x) = y$)。

**定义 14.11** 对偶

设 $\mathscr{C}$ 是一个范畴,以 $\mathscr{C}$ 的对象为对象,以 $\mathscr{C}$ 的态射的反向为态射形成一个新范畴,称为

$\mathscr{C}$ 的对偶，记作 $C^{op}$（即 $f\in C^{op}(A,B)$ 当且仅当 $f\in\mathscr{C}(B,A)$）。

对偶原理（Duality Principle）：设 $P$ 是一个关于所有范畴的真命题，则将命题 $P$ 中所有的态射反向得到的新命题 $P^*$ 也是一个关于所有范畴的真命题。对范畴论中的任意一个命题 $P$，我们有 $(P^*)^*=P$ 成立。由对偶原理可知，命题 $P$ 成立当且仅当其对偶命题 $P^*$ 成立，因此对于范畴论中的任意一对对偶命题，我们只需要证明其中一个命题成立，另一个即成立。

**定义 14.12**　同构

设 $A,B$ 是范畴 $\mathscr{C}$ 中的两个对象，$f:A\to B$，如果存在态射 $g:B\to A$ 使得

$$gf=1A,\quad fg=1B$$

则称态射 $f$ 是同构。如果存在同构 $f:A\to B$，则称 $A$ 与 $B$ 是同构的对象（Isomorphic Object）。容易验证同构是 ob$\mathscr{C}$ 上的一个等价关系。

在范畴论中，我们不仅关注对象，而且更关注对象之间的对应关系，即范畴之间的映射函子。

**定义 14.13**　函子

设 $\mathscr{C}$ 和 $\mathscr{D}$ 是范畴。函子 $F:\mathscr{C}\to\mathscr{D}$ 由两个映射组成：

$$\mathrm{ob}\mathscr{C}\to\mathrm{ob}\,\mathscr{D}:A\to F(A),$$

$$\mathrm{Mor}\mathscr{C}\to\mathrm{Mor}\,\mathscr{D}:f\to F(f)$$

满足 $\mathrm{dom}(F(f))=F(\mathrm{dom}(f))$，$\mathrm{cod}(F(f))=F(\mathrm{cod}(f))$，$F(1_A)=1_{F(A)}$，并且若 $\mathrm{dom}(g)=\mathrm{cod}(f)$，则 $F(gf)=F(g)F(f)$。

**定义 14.14**　反变（Contravariant）函子 $G$

将 $\mathscr{C}$ 中由 $A$ 到 $B$ 的态射 $f$ 送到 $\mathscr{D}$ 中由 $G(B)$ 到 $G(A)$ 的态射 $G(f)$ 的，并将 $F(gf)=F(g)F(f)$ 换为 $G(gf)=G(f)G(g)$，其他与上述定义相同。

$C^{op}$ 是一个范畴，具有与 $C$ 同样的对象；$C^{op}$ 中的态射仍是 $C$ 中的态射，但逆转了方向。如 $C^{op}$ 中态射 $f^{op}:B\to A$，即是 $C$ 中态射 $f:A\to B$。这样，$C$ 到 $D$ 的反变函子 $G$ 就是 $C^{op}$ 到 $D$ 的共变函子 $G:C^{op}\to D$。

函子是用来研究范畴之间的对应关系，自然变换是研究函子之间的对应关系。

**定义 14.15**　自然变换

设 $C$ 与 $D$ 是范畴，$F:C\to D$ 与 $G:C\to D$ 是两个函子，自然变换（Natural Transformation）$\alpha:F\to G$ 是一个映射 ob$C\to$ Mor$D$：

$$A\to(\alpha_A:F(A)\to G(A)),\quad A\in\mathrm{ob}C$$

使得对 $C$ 中的任意态射 $f:A\to B$，$G(f)\alpha_A=\alpha_B F(f)$ 成立，即图 14.1 中的变换。

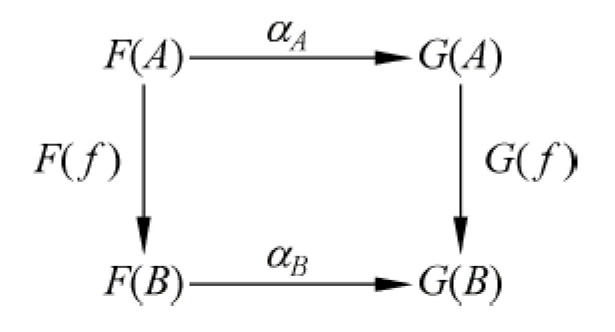



图 14.1　自然变换

如果自然变换 $\alpha:F\to G$ 满足对任意的 $A\in\mathrm{ob}C$，$\alpha_A:F(A)\to G(A)$ 是一个同构，则称 $\alpha$ 是一个自然同构（Natural Isomorphism）。

## 14.6 Topos

20 世纪 60 年代早期,格罗滕迪克(Grothendieck)用希腊词 Topos(拓扑斯)表示数学对象的通用框架,提出用拓扑空间 $X$ 上的集值层(Set Valued Sheaf)的全体做成的范畴 Sh(X)作为推广了的拓扑空间 $X$,用以研究空间 $X$ 上的上同调。他把拓扑的概念推广到小范畴(Small Category) $C$ 上,称为一个景(Site)(或称为 Grothendieck 拓扑)。

劳维尔(Francis W Lawvere)研究了 Grothendieck topos 和布尔值模型构成的范畴,发现它们都具有真值对象 D。1969 年夏,劳维尔和蒂尔尼(Tierney)决定合作研究层论(Sheaf Theory)的公理化问题。20 世纪 70 年代初,他们发现了一个比层(Sheaf)更广的类可以用一阶逻辑来刻划,同时这也是一个泛化了的集合论,他们提出了初级 Topos(Elementary Topos)的概念。这样,Sh($X$),Sh($C$,$J$)以及布尔值模型构成的范畴是初级 Topos,但后者还包括了在层(Sheaf)之外的其他范畴。初级 Topos 同时具有几何和逻辑的特性。Topos 的核心思想是用连续变化的集合来代替传统的不变的常量的集合,为研究可变结构(Variable Structure)提供一个更为有效的基础。

### 14.6.1 Topos 的定义

**定义 14.16** Topos(拓扑斯)或者初级 Topos 是满足下列等价条件之一的范畴:

(1) 具有指数和子对象分类的完全范畴。

(2) 具有子对象分类和它的幂对象的完全范畴。

(3) 具有等价类和子对象分类的笛卡儿闭范畴。

1969 年,劳维尔和蒂尔尼最初给出了上述初级 Topos 的定义,具有完备性。Topos 不仅都有有限范围,而且都有有限的上极限,以及子对象分类、指数和幂对象。这意味着,Topos 范畴特别是终端对象、等价类、回调、所有其他限制、指数对象和子对象分类。

一般来说,终端对象允许我们考虑对象的全局元素(全局部分);子对象分类 Ω 让我们考虑子对象,正如我们将看到的断言广义真值;指数可以考虑从一个对象到所有其他对象的箭头。结合幂对象的笛卡儿封闭性和存在性,这些特性允许在 Topos 对象中处理箭头,某种程度上与范畴集合的常见属性非常相似。所以,在集合范畴中真正需要做的任何事情都可以在任何 Topos 中完成。

### 14.6.2 Topos 之间的态射

(1) 逻辑函子 $E$,$F$ 是两个 Topos,一个函子 $j$: $E \rightarrow F$ 称为逻辑的,如果 $f$ 保持有限极限、函数空间以及真值对象 Ω。例如,inclusion 函子 $S_f \hookleftarrow S$ 就是逻辑函子。

(2) 几何态射:一个从 $F$ 到 $E$ 的几何态射 $f$ 由两个函子($f^*$,$f_*$)组成,这里 $f^*$:$E \rightarrow F$,$f_*$:$F \rightarrow E$,并且 $f^*$ 是 $f_*$ 的 left adjoint,$f^*$ 保持有限极限。

如果 $f$($f^*$,$f_*$),$g$($g^*$,$g_*$)是两个从 $F$ 到 $E$ 的几何态射,一个自然变换 $\tau$:$f \rightarrow g$ 是一个从 $f^*$ 到 $g_*$ 的自然变换,或者是从 $g_*$ 到 $f_*$ 的自然变换。这样,Topos 几何态射以及自然变换构成一个二维范畴,即对每两个 Topos $F$ 和 $E$,$F$ 和 $E$ 之间的所有几何态射以

及它们之间的自然变换也构成一个范畴。

几何态射的名称来自层论。如果 $X$ 和 $Y$ 是两个拓扑空间，$f$ 是 $X$ 到 $Y$ 的一个连续映射，那么 $f$ 导致 $f_*$：$\mathrm{Sh}(X)\to\mathrm{Sh}(Y)$ 以及 $f^*$：$\mathrm{Sh}(Y)\to\mathrm{Sh}(X)$，它们构成一个几何态射。

## 14.6.3　Sheaf 理论

Sheaf 理论（层论）源于代数拓扑，是作为构造上同调（Cohomology）的工具，但是随着对它的进一步研究，sheaf 思想开始影响到不同数字领域，如微分几何、代数几何及逻辑等。Sheaf 理论是一种数学术语，提供从局部到整体的一个有力工具，层论作为一个理论，其基本内容是层系数上同调论，这正好为流形上的整体分析提供了强有力的工具。

考虑一个小范畴 2，如图 14.2 所示。

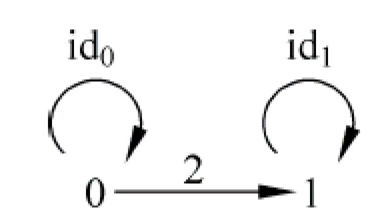



图 14.2　小范畴 2

有 2 个要素 1：$\mathrm{id}_1\in_1 1$、$2\in_0 1$ 和 1 个要素 0：$\mathrm{id}_0\in 0$。在范畴论中，我们可以问要素 1 什么时间（属于哪个阶段）属于要素 0，而不是全局集合论问题。这个问题更为普遍，因为它回答不需要集合 $\{0,1\}\cong\{\mathrm{true},\mathrm{false}\}$。答案可以是部分为真。图 14.3 给出了范畴 2 对象的全部 sheaf。

| object | sheafs on it |
| --- | --- |
| 0 | $\{\mathrm{id}_0\}=\uparrow 0$ |
| 0 | $\varnothing$ |
| 1 | $\{\{\mathrm{id}_1\},\{2\}\}=\uparrow 1$ |
| 1 | $\{2\}$ |
| 1 | $\varnothing$ |

图 14.3　范畴 2 对象的全部 sheaf

$X$ 是一个拓扑空间，$T$ 是所有开集构成的集合。$T$ 是一个偏序集，因而 $T$ 是一个范畴，具有末端对象 $X$，始端对象 $\varnothing$（空集）。

**定义 14.17**　Presheaf

$P$ 是 $T$ 上的一个 presheaf，如果 $P$ 是 $T$ 到 $S$ 的一个反变函子，即对任一开集 $U$，$P(U)$ 是一个集合。若 $U\leqslant V$，那么有一个集合映射 $P_U^V$：$P(V)\to P(U)$ 如果 $U\leqslant V\leqslant W$，那么 $P_U^V\circ P_V^W=P_U^W$。

**定义 14.18**　Sheaf

$P$ 是 $T$ 上的一个 presheaf，$\{Ua\}_{a\in A}$ 是任一开集 $U$ 的任一开覆盖，如果一族 $\{a_\alpha\in P(U_a)\,|\,a\in A\}$ 具有性质：对任意 $\alpha,\beta\in A$，$P_{U_\alpha^\beta}^{U_\alpha}u_\beta(a_\alpha)=P_{U_\alpha^\beta}^{U_\beta}u_\beta(a_\beta)$ 成立（具有此性质的一族 $\{a_\alpha\}$ 称为一个相容族（Compatible Family），则存在唯一 $\alpha\in P(U)$，使得 $P_{U_\alpha}^U(\alpha)=a_\alpha$ 任意 $\alpha\in A$，那么称 $P$ 为一个 Sheaf。

**定义 14.19**　Grothendieck 拓扑

$C$ 是一个小范畴，对 $C$ 的任一对象 $c$，$J(c)$ 是 $c$ 上的一些 sieve 的集合（$c$ 上的一个 sieve，即 $c$ 在 $S^{C\mathrm{op}}$ 中的一个子对象），称为 $c$ 的一个覆盖 sieve。$J$ 是 $C$ 上的一个 Grothendieck 拓扑，如果 $f$ 满足以下条件：

(1) 对任意 $c$，极大 sieve$\{a\,|\,\mathrm{codomain}(a)=c\}\in J(c)$。

(2) $f: c' \to c$ 是 $C$ 中任一以 $c$ 为 codomain 的态射,如果 $R \in J(c)$,那么

$$f^*(R) = \{c'' \beta \to c' \mid f \circ \beta \in R\} \in J(c')$$

这里图 14.4 是 $S^{C^{op}}$ 中的 pullback 图形。

(3) $R \in J(c)$,$s$ 是 $c$ 上的一个 sieve,如果对 $R$ 中任意 $f: c' \to c$,$f^*(s) \in J(c')$,那么 $s \in J(c)$。

$J$ 是 $C$ 上的一个 Grothendieck 拓扑,$F$ 是 $C$ 上的一个 presheaf,即 $F$ 是 $S^{C^{op}}$ 中的一个对象。如果对任意 $R \in J(c)$,$\tau$ 是 $R \to F$ 在 $S^{C^{op}}$ 中的一个态射。那么存在唯一 $S^{C^{op}}$ 中态射 $\bar{\tau}$: $h_c \to F$,使得三角形 $\begin{array}{ccc} R & \hookrightarrow & h_c \\ & {}_{\tau}\searrow \quad \swarrow_{\bar{\tau}} & \\ & F & \end{array}$ 在 $S^{C^{op}}$ 中可换,即 $\bar{\tau}$ 是 $\tau$ 的一个扩张。这样称 $F$ 是 $C$ 上的一 $J$-sheaf。

记 Sh($C$,$J$)为由 $C$ 上 $J$-sheaf 以及它们之间的自然变换构成的 $S^{C^{op}}$ 的子范畴。$Sh(C$,$J)$是一个 topos。

$$Sh(C, J) \underset{i}{\overset{L}{\rightleftarrows}} S^{C^{op}}$$

inclusion 函子 $i$ 的 left adjoint $L$ 称为 sheaf 化函子(Sheafification Functor)。$L$ 保持有限极限,所以($i$,$L$)构成一个从 Sh($C$,$J$)到 $S^{C^{op}}$ 的几何态射。

## 14.6.4 Topos 的内逻辑

1973 年,迪亚科尼斯库(Diaconescu)提出了内范畴的概念,进而开始了对 Topos 内结构的研究。在某种意义上说,Topos 是一个推广了的集合论。

一个小范畴 $C$,令 $C_0$ 表示 $C$ 中全体对象,$C_1$ 表示 $C$ 中的全体态射,则 $C_0$,$C_1 \in S$,即 $C_0$,$C_1$ 是集合。描述 $C$ 是一个范畴的公理就可以重新用 $S$ 中的某些映射及交换图形来代替。例如每一对象 $c$ 有一个单位元(恒等映射) $1c$,即存在映射($S$ 中态射) $i$: $C_0 \to C_1$, $c \to 1c$。对任意 $f \in C_1$, domain 和 codomain 是两个映射,$d_0$,$d_1$: $C_1 \to C_0$,$d_0(f) = \mathrm{dom}(f)$,$d_1(f) = \mathrm{cod}(f)$。映射的结合描述为图 14.5 所示的 $S$ 中的 pullback 图形。




图 14.4 $S^{C^{op}}$ 中的 pullback 图形　　图 14.5 S 中的 pullback 图形

这里 $C_1 \times_{C_0} C_1 = \{(f,g) \mid d_0(g) = d_1(f)\}$是 $C_1$ 中的可结合映射对。映射的结合 $m$: $C_1 \times_{c_0} C_l \to C_1$,$rn(f,g) = g \circ f$。$i$,$d_0$,$d_1$,$m$ 需要满足 $d_0 \circ i = d_1 \circ i = 1_{c0}$; $d_0 \circ m = d_0$。$\Pi_1$,$d_1 \circ m = d_1 \circ \Pi_2$,$m \circ (1_{c1} \times m) = m \circ (m \times 1_{c1})$,$m \circ (1_{c1} \times i) = m \circ (i \times 1_{c1}) = 1_{c1}$。

以上给出了有有限极限的范畴 $E$ 的内范畴的定义。同样,从内范畴 $C$ 到 $D$ 的内函子 $F$,可以定义为 $E$ 中两个态射 $F_0$: $C_0 \to D_0$,$F_1$: $C_1 \to D_1$,满足某些条件(即反映 $F(g \circ f) = F(g) \circ F(f)$及 $F(1_c) = 1_{F(C)}$ 的可换图形)。

范畴 $S$ 中的逻辑是建立在元素的从属关系上的。在 Topos $E$ 中,元素以及逻辑运算等都成为态射。若 $X$ 是一个集合,$X$ 的一个元素 $x(x \in X)$,可以看作一个映射 $x$: $1 \to X$。

这样，可以定义在有末端对象 1 的范畴 $E$ 中任一对象 $X$ 的一个元素为一个从 1 到 $X$ 的态射。例如，取 $M$ 是一个 monoid，$E = S^{M^{op}}$ 是右 $M$-集的范畴，$X$ 是 $E$ 中一个对象，即 $X$ 是一个右 $M$-集，一个映射 $1 \to X$ 则是右 $M$-集 $X$ 的一个固定点(Fixed Point)。如果 $X$ 没有固定点，则 $X$ 没有元素。所以这样定义的元素不能确定所给的对象。元素的概念必须加以推广。

**定义 14.20**　广义元素(Generalized Element)

$E$ 中对象 $X$ 的一个广义元素是以 $X$ 为 codomain 的 $E$ 中任一态射。

这样，通常基于元素从属关系的集合论的一些概念、语言等就可以用到 Topos 中来。一般来说，一个 Topos 的内逻辑是直觉类型理论(Intuitive Type Theory)。

在 Topos $E$ 中，对每一个对象 $A$，可以结合变量 $a$，$a'$，$a''$，…。一个具有自由变量 $a_1,\cdots,a_n$，type $A$ 的项是一个态射 $A_1\times\cdots\times A_n \to A$，这里 $A_i(i=1,\cdots,n)$ 是 $A_i$ 的 type，常记为 $a_i: A_i$($a_i$ 具有 type $A_i$)。

一个公式或 propositional function $\varphi$ 是一个具有 type$\Omega$ 的项。这样，如果 $\varphi$ 具有自由变量 $a_1,\cdots,a_n$，那么 $\varphi$ 的解释是 $A_1\times\cdots\times A_n$ 的子对象 $\{A_1\times\cdots\times A_n|\varphi\}$(有时记为 $\{(a_1,\cdots,a_n)\in A_1\times\cdots\times A_n|\varphi(a_1,\cdots,a_n)\}$)，它具有特征函数 $\varphi$。称 $\varphi$ 为普效的，如果 $\{A_1\times\cdots\times A_n|\varphi\}=A_1,\cdots,A_n$。

逻辑运算是 $\Omega$ 之间的态射。

$\neg$: $\Omega\to\Omega$ 是 $0 \rightarrowtail \Omega$ (0 是始端对象)的特征函数。

$\wedge$: $\Omega\times\Omega\to\Omega$ 是 $(t,t)$: $1\rightarrowtail\Omega\times\Omega$ 的特征函数。

$\Rightarrow$: $\Omega\times\Omega\to\Omega$: 先取 equalizer $\Omega_1 \rightarrowtail \Omega\times\Omega \overset{A}{\underset{n_2}{\rightrightarrows}} \Omega$，$\Omega_1$ 是一个顺序关系，“$a\leqslant b$ 当且仅当 $a\wedge b=a$”。这样，$\Omega_1$ 是 $\Omega\times\Omega$ 一个子对象，它的特征函数就是$\Rightarrow$: $\Omega\times\Omega\to\Omega$。“$\Rightarrow$”的解释是这样的：令 $\varphi_1,\varphi_2$ 是两个公式，$\varphi_1,\varphi_2$ 分别是 $X_1\rightarrowtail X$ 和 $X_2\rightarrowtail X$ 的特征函数。那么 $\varphi_1\Rightarrow\varphi_2$ 是 $Y\rightarrowtail X$ 的特征函数，这里 $Y$ 是 $X$ 的最大的具有性质 $Y\cap X_1\leqslant X_2$ 的子对象。

$\varphi_1,\varphi_2$ 是两个公式，假设都具有自由变量 $a_1,\cdots,a_n$，则 $A_1\times\cdots\times A_n$ 的子对象 $\{A_1\times\cdots\times A_n|\varphi\wedge\psi\}$ 的特征函数为 $\wedge\circ(\varphi_1,\varphi_2)$: $A_1\times\cdots\times A_n \xrightarrow{(\varphi_2,\ \circ\varphi_1)} \Omega\times\Omega \xrightarrow{\wedge} \Omega$。$\wedge\circ(\varphi_1,\varphi_2)$ 常记为 $\varphi_1\wedge\varphi_2$。同样 $\varphi_1\Rightarrow\varphi_2$ 记为$\Rightarrow\circ(\varphi_1,\varphi_2)$。这样，施行逻辑运算成为对态射的结合。

令 $X,Y$ 是 $E$ 的两个对象，$P$: $X\times Y\to Y$ 是从 $X\times Y$ 到 $Y$ 的射影。由 topos 基本定理，$P$ 导致一个 essential 几何态射。由 $Y$ 或 $X\times Y$ 的子对象构成的范畴 Sub($Y$) 或 Sub($X\times Y$) 分别是 $E/Y$ 或 $E/X\times Y$ 的子范畴。因为 $\Pi_p$ 保持有限极限，特别末端对象以及 monomorphism，所以 $\Pi_p$ 在 Sub($X\times Y$) 上的限制给我们一个从 Sub($X\times Y$) 到 Sub($Y$) 的函 $\forall_x$。因为 $P^*$ 保持有限极限，所以 $P^*$ 在 Sub($Y$) 上的限制(仍用 $P^*$ 表示)是 Sub($Y$) 到 Sub($X\times Y$) 的函子 $\forall_x$。

我们有 $\exists_x$ 是 $P^*$ 的 left adjoint，$P^*$ 是 $\forall_x$ 的 left adjoint，如图 14.6 所示。如果 $A$ 是 $X\times Y$ 的一个子对象(子集)，$B$ 是 $Y$ 的一个子对象(子集)，则

$$\exists_x A\leqslant B \text{ 当且仅当 } A\leqslant P^*B,$$
$$B\leqslant\forall_x A \text{ 当且仅当 } P^*B\leqslant A$$

Sub($X\times Y$) $\overset{\exists_X}{\longrightarrow}$ / $\overset{P^*}{\longleftarrow}$ / $\overset{\forall_X}{\longrightarrow}$ Sub($Y$)

图 14.6　量词

内逻辑的形式构造法有很多种，这里采用布瓦洛(A Boileau)和乔亚尔(A Joyal)提出的

Topos 内部语言[90]。$E$ 是一个 Topos,语言 $L_E$ 由以下成分组成:

(1) types。$A,B,\cdots$是 $E$ 中的对象。

(2) 对任一 type $A$ ,项的定义如下:

① 可数的,以 $A$ 为 type 的变量 $a,a',\cdots$。

② 如果 $f: A\to B$ 是$E$ 中一个态射,$t: A$ 是一个项,那么 $f(t)$是一个具有 type $B$ 的项,$f(t): B$。

③ 如果 $R\subseteq A_1\times\cdots\times A_n$ 是一个关系(即 $R$ 是 $A_1\times\cdots\times A_n$ 的一个子对象),并且 $t_1: A_1,\cdots,t_n: An$,那么 $R(t_1,\cdots,t_n): \Omega$。

④ 如果 $r: A,s: B$ ,那么$<r,s>: A\times B$。

⑤ 如果 $t: A\times B$ ,那么$\Pi_1(t): \Pi_2(t): B$。

⑥ 如果 $r,s: A$,那么$(r=s): \Omega$。

⑦ $T: Q$ ($T$ 表示真)。

⑧ 如果 $\varphi,\psi: \Omega$,那么$(\varphi\wedge\psi): \Omega$。

⑨ 如果 $t: B^A,r: A$,那么$(t`r): B$ (若 $B=\Omega$,常记 $t`r$ 为$r\in t$,即认为 $t$ 是$A$ 的一个“子集”。)

⑩ 如果 $s: B$ ,$a: A$ ,那么$(\lambda\alpha\cdot.s): B^A$。

(3) 公式即具有 type $\Omega$ 的项。

$\Gamma=\{r_1,\cdots,r_n\}$为有限个公式,$\varphi$是一个公式。用$\wedge\Gamma$表示公式 $r_1\wedge\cdots\wedge r_n$。$V=\{x_1,\cdots,x_m\}$是出现在 $r_1,\cdots,r_n,\varphi$ 中的自由变量,$x_i: X_i,i=1,\cdots,m$。式子 $\Gamma|\bar{V}^{\varphi}$ 称为一个 sequent。如果$\{X_1\times\cdots\times X_m|\Gamma|\bar{V}\varphi\wedge\Gamma\}\leqslant\{X_1\times\cdots\times X_m|\varphi\}$($\leqslant$是偏序集 $\mathrm{Sub}(X_1\times\cdots\times X_m)$中的顺序),那么我们说 sequent 是普效的。

### 14.6.5 公理和推理

(1) $\varphi|\overline{\;V\;}\varphi$。

(2) $\dfrac{\Gamma|\overline{\;V\;}\varphi}{\Gamma,\Pi|\overline{\;V,U\;}\varphi}$。

(3) $\dfrac{\Gamma|\overline{\;U\;}\varphi;\ \Pi,\varphi|\overline{\;V\;}\psi}{\Gamma,\Pi|\overline{\;U,V\;}\psi}$。

(4) $\Gamma|\overline{\;V\;}T$。

(5) $\dfrac{\Gamma|\overline{\;U\;}\varphi;\ \Pi|\overline{\;V\;}\psi}{\Gamma,\Pi|\overline{\;U,V\;}\varphi\wedge\psi}$;$\dfrac{\Gamma|\overline{\;V\;}\varphi\wedge\psi}{\Gamma|\overline{\;V\;}\varphi}$;$\dfrac{\Gamma|\overline{\;V\;}\varphi\wedge\psi}{\Gamma|\overline{\;V\;}\psi}$。

(6) $|\overline{\;V\;}t=t$(或者 $T|\overline{\;V\;}t=t$)。

(7) $\dfrac{\Gamma \mid \overline{\ \ U\ \ }\varphi(s)\text{；}\Pi \mid \overline{\ \ V\ \ }t=t}{\Gamma,\Pi \mid \overline{U,V}\varphi(t)}$。

(8) $\mid \overline{\ \ V\ \ }(\lambda x \cdot t(x))`r=t(r)$。

(9) $\dfrac{\Gamma,\varphi \mid \overline{\ \ U\ \ }\psi\text{；}\Pi,\psi \mid \overline{\ \ V\ \ }\varphi}{\Gamma,\Pi \mid \overline{U,V}\varphi=\psi}$。

(10) $\mid \overline{\ \ \ }\Pi_1<t,s>=t$；$\mid \overline{\ \ \ }\Pi_2<t,s>=s$；$\mid \overline{\ \ \ }<\Pi_1 s,\Pi_2 s>=s$。

(11) $\dfrac{\Gamma \mid \overline{V,x}t=r}{\Gamma \mid \overline{\ \ V\ \ }(\lambda x \cdot t)=(\lambda x \cdot r)}$ • $x$ 不在 $\Gamma$ 中自由出现。

通常的逻辑运算可以定义如下：

(1) $(\forall x\varphi)\equiv(\lambda x \cdot \varphi)=(\lambda x \cdot T)$，$x$：$B$。

(2) $(\varphi\Rightarrow\psi)\equiv(\varphi\wedge\psi=\varphi)$。

(3) $(\varphi\vee\psi)\equiv\forall p[((\varphi\Rightarrow p)\wedge(\psi\Rightarrow p))\Rightarrow p]$，$p$：$\Omega$。

(4) $\perp\equiv\forall p(p)$，$p$：$\Omega$($\perp$：假)。

(5) $(\exists x\varphi)\equiv\forall p[\forall x(\varphi\Rightarrow p)\Rightarrow p]$，$x$：$B$，$p$：$\Omega$。

关于这些运算的通常规律可以从以上定义以及公理和推理规则推出。

## 14.7　心理逻辑

数理逻辑始于莱布尼茨(G W Leibniz)，在布尔和弗雷格处发生了分流，形成了所谓的逻辑的代数传统和逻辑的语言传统。在图灵机理论中，图灵核心阐述了“自动机”和“指令表语言”这两个概念，这两者很好地契合了莱布尼茨关于“理性演算”和“普遍语言”的构想。皮亚杰在儿童思维产生及其发展的研究过程中，发现了心理运算的结构，改造经典的数理逻辑，创立了一种新型的逻辑——心理逻辑(Psycho-logic)，并用来描述儿童不同智力水平的认知结构[877]。这种逻辑包括具体运算和形式运算两个系统。具体运算主要有类和关系的八个群集，形式运算则主要包括 16 种命题的运算以及 INRC 群结构。皮亚杰的心理逻辑系统更新了我们对逻辑的观念，成为解决逻辑认识论问题的基础，用逻辑结构来刻画认知结构。晚年皮亚杰在《走向意义的逻辑》《态射与范畴》《可能性与必然性》等一系列新著作中，以一种更新的、更有力的方式去修正和发展他的理论，以至贝林(H Beilin)将其称为“皮亚杰的新理论”[67]。

### 14.7.1　组合系统

皮亚杰认为，当儿童思维可以脱离具体事物进行时，其首要成果便是使事物间的“关系”和“分类”从它们具体的或直觉的束缚中解放出来，组合系统使儿童的思维能力得到了扩展

和增强。所谓十六种二元命题，一般称为含有二个支命题的复合命题可能具有的十六种类型的真值函项”，表 14.2 给出了二元复合命题的十六种类型真值函项。

**表 14.2　二元复合命题的十六种类型真值**

| $(p,q)$ | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | $f_{12}$ | $f_{13}$ | $f_{14}$ | $f_{15}$ | $f_{16}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1,1) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| (1,0) | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| (0,1) | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| (0,0) | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

一般数理逻辑书中，以 $p\vee q$，$p\rightarrow q$，$p\leftrightarrow q$，和 $p\wedge q$ 这四个最基本的二元真值形式，即析取式、蕴涵式、等值式和合取式来分别表示 $f_2$，$f_5$，$f_7$，$f_8$ 这四种真值函项。皮亚杰对其余的命题函项也加以命名：$f_1$ 为 $p\cdot q$（完全肯定），$f_3$ 为 $p\leftarrow q$（反蕴涵），$f_4$ 为 $p(q)$（$p$ 的肯定、$f_9$ 为 $p/q[\bar{p}\vee\bar{q}]$（不相容）、$f_{10}$ 为 $pw$（互反排斥）、$f_{11}$ 为 $q[\bar{p}]$（$q$ 的否定）、$f_{12}$ 为 $p\cdot\bar{q}$（非蕴涵）、$f_{13}$ 为 $\bar{p}[q]$（$p$ 的否定）、$f_{14}$ 为 $q\cdot\bar{p}$（非反蕴涵）、$f_{15}$ 为 $\bar{p}\cdot\bar{q}$（合取否定；非析取）、$f_{16}$ 为(0)完全否定。皮亚杰认为它们体现于青少年的实际思维之中，构成他们的认知结构。

### 14.7.2　INRC 四元群结构

INRC 转换群是形式思维出现的另一种认知结构，它与命题运算关系密切。皮亚杰以两种可逆性，即反演和互反为轴，将它们构成四种不同类型的 INRC 转换群。皮亚杰试图以此为工具，阐明现实的思维机制，特别是它的可逆性质。以可逆性概念贯穿于分析主体的智慧发展过程，这是皮亚杰理论的特色之一。

INRC 四元群的含义是：任何一个命题都有相应的四个转换命题，或者说，它可以转换成四个互相区别的命题。其中有一个转换是重复原来的命题(I)，称为恒等性转换。另外三个转换是依据反演可逆性的反演转换(N)、依据互反可逆性的互反性转换(R)以及建立在这两种可逆性基础之上的对射性转换(C)。这四种转换所生成的四个命题(其中有一个是原命题)就构成了一个关于“转换”的群。虽然只有四个命题，即四个元，但它们之间的关系符合群结构的四个基本条件。四元转换群中两种可逆性的综合体现在对射性转换上，因为对射就是互反的反演或反演的互反，即 C=NR 或 C=RN。

由此可知，四元转换群实质就是二元复合命题通过算符(如合取、析取、蕴涵等)之间的内在联系而形成的某种整体组织。因此，分析四元群结构不能不从命题出发。皮亚杰认为，十六种二元命题构成了四种类型的四元转换群。

A 型：析取、合取否定、不相容和合取构成 A 型四元群。

B 型：蕴涵、非蕴涵、反蕴涵和非反蕴涵构成 B 型四元群。

C 型和 D 型是两种特殊型，在 C 型中，原运算与互反运算相同；反演运算与对射运算相同。“完全肯定”与“完全否定”，“等价”与“互相排斥”构成 C 型的两个亚型。在 D 型中，原运算与对射运算相同；反演运算与互反运算相同。“p 的肯定”与“p 的否定”，“q 的肯定”与“q 的否定”构成 D 型的两个亚型。

INRC 的集合具有以下性质。

(1) 集合中的两个元素的组合仍是集合内的一个元素(封闭性)。

(2) 组合是结合性。

(3) 每一个元素有一个逆运算。

(4) 有一个中性元素(I)。

(5) 组合是可交换的。

## 14.7.3 态射—范畴论

皮亚杰的形式化工作可区分为两个阶段:早期的结构主义时期和晚年的后结构主义时期。前者又称为经典理论,后者称为新理论阶段。皮亚杰的新形式化理论基本上放弃了运算结构论,而代之以态射-范畴论。于是传统的前运算-具体运算-形式运算的发展系列变成了内态射(Intramorphic)—间态射(Intermorphic)—超态射(Extramorphic)的发展系列。

第一阶段称为内态射水平。心理上只是简单的对应,没有组合。共有的特点都是基于正确的或不正确的观察,特别是以可见的预测为基础。这仅是一个经验的比较,依赖简单的状态转换。

第二阶段称为间态射水平,标志着系统性的组合建构开始。间态射水平的组合建构只是局部的、逐步发生的,最后并没有建构成一个封闭性的一般系统。

最后阶段是超态射,主体借助运算工具进行态射的比较。而其中的运算工具,正是对组成先前态射内容进行解释和概括而得到。

皮亚杰采用了如图 14.7 所示的实验装置[573]。这套装置由直径不同的圆盘组成,其中心钉于一个支架上。每个圆盘的顶端都有一个能挂上不同重物的砝码栓,挂上重物后圆盘会向不同方向旋转。主试要求儿童向两个或更多的圆盘上挂重物,但不能使圆盘旋转,也就是说要保持平衡。

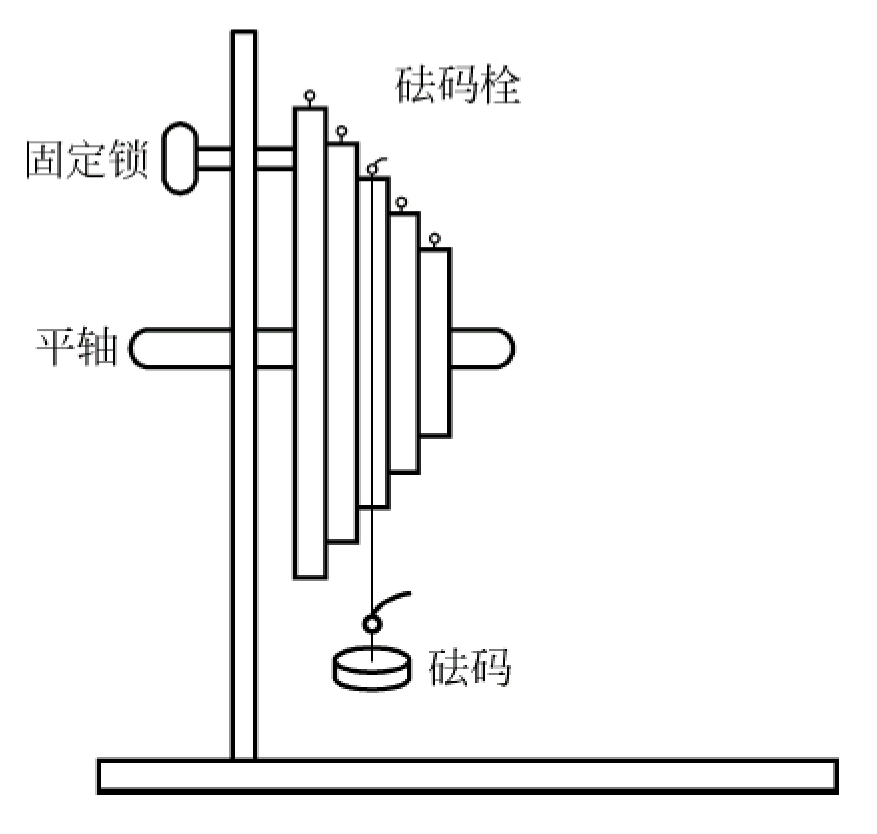


图 14.7 同轴盘装置

第一组实验观察到,即使是能圆满完成标准守恒任务的 9 岁半儿童,也仅仅是以简单的对应性为基础进行推理:重量的大小↔影响的大小。这种推理称为“心理内态射推理”,此时儿童以对应性中的共变为基础进行预测。

第二组 11、12 岁的儿童开始运用第二种对应性,即圆盘的大小↔影响的大小。这一组的儿童能够意识到,在大盘上同样的重物会比在小盘上产生更大的力量,也就是说,他们开始将重物和圆盘这两种对应性联合起来考虑,这时儿童就达到了“心理间态射水平”。但问题在于,处于心理间态射水平的儿童并不能解释圆盘保持平衡的原因,他们还无法判断出这两种对应性之间的相互依赖关系,即重物产生力的大小依赖于圆盘的大小,它们之间是一种交互的或相乘的关系。此时儿童仅根据可见的关系(对应性)进行推理,还不能凭借抽象的、不可见的关系进行推理。

要想达到新的态射水平,必须涉及一种概括化过程,即从经验的对应性发展到以转换为基础的、更抽象的对应性。例如,具有超态射水平的儿童可以意识到,如果增加更多的苹果,那么将得到一个更大的水果集合,即对水果的转换也会同时发生。这样,发展就成了将转换

变为在层级水平上的整体对应性问题,它超越了经验式的对应性。儿童一般到 12～15 岁才能达到超态射水平。

## 14.8 认知动力学

1995 年,冯·盖尔德(T van Gelder)和波特(R Port)出版了一本关于认知科学的动力理论的书,提出认知科学的动力学研究之路[778]。动力学假说是以数学的动力系统理论为基础描述认知的,用数学中的状态空间、吸引子、轨迹、确定性混沌等概念来解释与环境相互作用的认知智能体的内在认知过程。用微分方程组来表达处在状态空间的认知智能体的认知轨迹。换句话说,认知是作为认知智能体所有可能的思想和行为构成的多维空间来描述的,特别是通过在一定环境下和一定的内部压力下的认知智能的思想轨迹来详尽考察认知的。认知智能的思想和行为都受微分方程的支配。系统中的变量是不断进化的,系统服从于非线性微分方程,一般来讲是复杂的、确定性的。

认知动力学强调多水平分析的动态综合,认为所有的发展结果都可以解释为通过系统较简单的成分之间循环的交互作用,而自发出现的连贯的较高等级的形式。这种过程就是所谓的自组织。动力系统观用自适应组织和自适应稳定来解释源自内在过程的等级的产生、复杂性的提高、发展系统内新异性的出现、过渡阶段多发生的结构变化和个体多样化以及对环境的适应能力等。

动力论的认知范式与其他范式的一个重要区别是对表征的不同理解[893]。符号主义模型是以符号表征为基础的。联结主义的表征是以网络中的并行式表征或局部符号表征为基础的。但动力论的认知范式则宣称,一个动力模型应当是"无表征的"。动力系统理论对认知行为的连续性提供了随时间变化的自然主义的说明。动力系统理论的优势是对认知的描述是多元的,是一种经验可检验的理论,可以对描述认知系统的微分方程进行分析修正,也可以用已知的技术去解这些方程,比起其他理论,它是一种定量的分析,是理解认知的一种确定性的观点。另一优势是动力系统的描述可以展示人类行为复杂的、混沌的特性。

对于认知动力系统表征的理解受到质疑。如何保证动力系统的各变量和参数的恰当选择、如何保证系统的稳定性和可靠性等问题都需要进一步研究。

# 第 15 章<<

# 智能机器人

智能机器人拥有相当发达的“人工大脑”，可以按目的安排动作，还具有传感器和效应器。机器人技术的发展是一个国家高科技水平和工业自动化程度的重要标志和体现。在 20 世纪末计算机文化已深入人心的基础上，机器人文化将在 21 世纪对社会生产力的发展，对人类生活、工作、思维的方式以及社会发展产生无可估量的影响。

## 15.1 概述

机器人的诞生和机器人学的建立及发展，是 20 世纪自动控制领域最具说服力的成就，是 20 世纪人类科学技术进步的重大成果。目前，许多国家都已经把机器人技术列入本国 21 世纪高科技发展计划，各种机器人系统向具有更高智能和与人类社会融洽更密切的方向发展。人机融合和人机共存，多样化的人类与机器配合，共同解决问题，是智能机器人发展的方向。

机器人的起源可追溯到 3000 多年前。早在我国西周时代(公元前 1066 年至前 771 年)，就流传着有关巧匠偃师献给周穆王一个艺妓(歌舞机器人)的故事。春秋时代(公元前 770 至前 467)后期，被称为木匠祖师爷的鲁班，利用竹子和木料制造出一个木鸟，它能在空中飞行。1893 年，加拿大摩尔设计的能行走的机器人“安德罗丁”，是以蒸汽为动力的。

1920 年，捷克作家卡佩克(Karel Capek)发表了科幻剧本《罗萨姆的万能机器人》。在剧本中，卡佩克把捷克语 Robota 写成了 Robot，Robota 是奴隶的意思。该剧预告了机器人的发展对人类社会的悲剧性影响，引起了大家的广泛关注，被当成了“机器人”一词的起源。

卡佩克提出了机器人的安全、感知和自我繁殖问题。针对人类社会对即将问世的机器人的不安，美国著名科学幻想小说家阿西莫夫于 1950 年在他的小说《我是机器人》中，首先使用了机器人学(Robotics)这个词来描述与机器人有关的科学，并提出了有名的“机器人三守则”。

(1) 机器人必须不危害人类，也不允许它眼看人将受害而袖手旁观。

(2) 机器人必须绝对服从于人类，除非这种服从有害于人类。

(3) 机器人必须保护自身不受伤害，除非为了保护人类或者是人类命令它做出牺牲。

这三条守则，给机器人社会赋以新的伦理性，并使机器人概念通俗化，更易于为人类社会所接受。至今，它仍为机器人研究人员、设计制造厂家和用户，提供了十分有意义的指导方针。

通常可将机器人分为三代。第一代是可编程机器人。这类机器人一般可以根据操作员所编的程序，完成一些简单的重复性操作。这一代机器人从 20 世纪 60 年代后半期开始投

人使用,目前他在工业界得到了广泛应用。第二代是感知机器人,即自适应机器人,它是在第一代机器人的基础上发展起来的,具有不同程度的“感知”能力。这类机器人在工业界已有应用。第三代机器人将具有识别、推理、规划和学习等智能机制,它可以把感知和行动智能化结合起来,因此能在非特定的环境下作业,故称为智能机器人。

1967年在日本召开的第一届机器人学术会议上,提出了两个有代表性的机器人定义。一个是森政弘(Masahiro Mori)与合田周平提出的“机器人是一种具有移动性、个体性、智能性、通用性、半机械半人性、自动性、奴隶性等7个特征的柔性机器”。从这一定义出发,森政弘又提出了用自动性、智能性、个体性、半机械半人性、作业性、通用性、信息性、柔性、有限性、移动性等10个特性来表示机器人的形象。另一个是加藤一郎提出的符合如下3个条件的机器称为机器人。

(1) 具有脑、手、脚等三要素的个体。

(2) 具有非接触传感器(用眼、耳接受远方信息)和接触传感器。

(3) 具有平衡觉和固有觉的传感器。

机器人的定义是多种多样的,其原因是它具有一定的模糊性。动物一般具有上述这些要素,所以在把机器人理解为仿人机器的同时,也可以广义地把机器人理解为仿动物的机器。

恩格尔伯格(Joseph F Engelberger)于1958年创建了世界上第一个机器人公司——Unimation(Univeral Automation)公司。1959年研制出了世界上第一台工业机器人,他被称为“机器人之父”。可以说,20世纪60年代和70年代是机器人发展最快、最好的时期,这期间的各项研究发明有效地推动了机器人技术的发展和推广。

机器人现在已被广泛地用于生产和生活的许多领域,按其拥有智能的水平可以分为三个层次。

(1) 工业机器人。它只能死板地按照人给它规定的程序工作,不管外界条件有何变化,自己都不能对程序也就是对所做的工作作相应的调整。如果要改变机器人所做的工作,必须由人对程序作相应的改变,因此它是毫无智能的。

(2) 初级智能机器人。它和工业机器人不一样,具有像人那样的感受、识别、推理和判断能力。可以根据外界条件的变化,在一定范围内自行修改程序,也就是它能适应外界条件变化对自己作相应调整。不过,修改程序的原则由人预先给以规定。这种初级智能机器人已拥有一定的智能,虽然还没有自动规划能力,但这种初级智能机器人也开始走向成熟,达到实用水平。

(3) 高级智能机器人。它和初级智能机器人一样,具有感觉、识别、推理和判断能力,同样可以根据外界条件的变化,在一定范围内自行修改程序。所不同的是,修改程序的原则不是由人规定的,而是机器人自己通过学习,总结经验来获得修改程序的原则。所以它的智能高出初级智能机器人。这种机器人已拥有一定的自动规划能力,能够自己安排自己的工作。这种机器人可以不要人的照料,完全独立的工作,故称为高级自律机器人。这种机器人已经走向实用。

自动驾驶汽车(Autonomous Vehicles; Self-piloting Automobile )又称无人驾驶汽车,是一种轮式移动机器人,是通过电脑系统实现无人驾驶的智能汽车。2010年10月9日,谷歌公司在官方博客中宣布,正在开发自动驾驶汽车。2015年6月11日,百度公司表示,百度与德国宝马汽车公司合作开发自动驾驶汽车。

## 15.2　智能机器人的体系结构

智能机器人体系结构指一个智能机器人系统中的智能、行为、信息、控制的时空分布模式。体系结构是机器人本体的物理框架，是机器人智能的逻辑载体，选择和确定合适的体系结构是机器人研究中最基础的并且非常关键的一个环节。以智能机器人系统的智能、行为、信息、控制的时空分布模式作为分类标准，沿时间线索归纳出 7 种典型结构：分层递阶结构、包容结构、三层结构、自组织结构、分布式结构、进化控制结构和社会机器人结构。

### 1. 分层递阶结构

1979 年，萨里迪斯（G Saridis）提出分层递阶结构，其分层原则是随着控制精度的增加而智能能力减少。他根据这一原则把智能控制系统分为 3 级，即组织级、协调级和执行级[641]。图 15.1 给出了分层递阶结构。

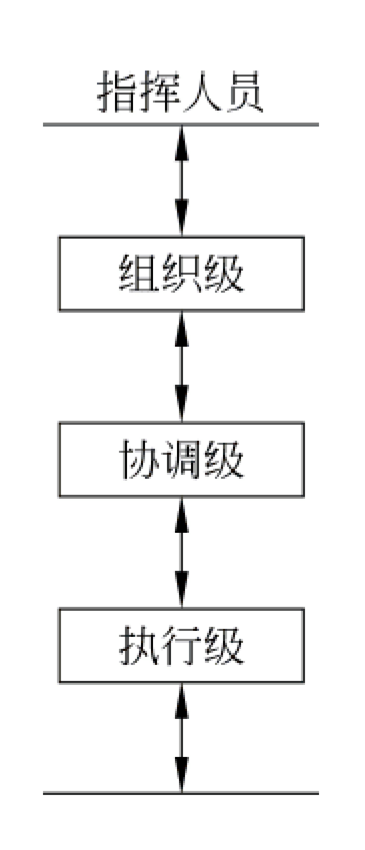



图 15.1　分层递阶结构

分层递阶结构是目标驱动的慎思结构，其核心在于基于符号的规划，其思想源于西蒙和纽尼尔的物理符号系统假说。分层递阶结构中两个典型的代表是 SPA（Sense-Plan-Act）和 NASREM。SPA 应用于第一个具有规划功能的移动机器人 Shakcy，该机器人控制系统划分为感知（S）、规划（P）、执行（A）3 个线性串联的模块。S 模块处理传感信息、环境建模；P 模块根据环境模型和任务目标进行规划；A 模块执行 P 模块规划结果。信息按 S—P—A 方向单向流动，无反馈。

美国航天航空局（NASA）和美国国家标准局（NBS）提出一个机器人 NASREM 结构体系[11]，它是一个严格按时间和功能划分模块的分层梯阶系统。系统对总命令一级一级进行时间和空间上的分解并根据需要调用传感器信息处理模块及相应的数据。NASREM 是 NASA/NBS 提出的参考模型并首先应用在空间机器人上，整个系统分成信息处理、环境建模、任务分解 3 列和坐标变换与伺服控制、动力学计算、基本运动、单体任务、成组任务、总任务 6 层，所有模块共享一个全局存储器（数据库），系统还包括一个人机接口模块，它是一个典型的、严格按时间和功能划分模块的分层递阶系统。

分层递阶结构智能分布在顶层，通过信息逐层向下流动，间接地控制行为。该结构具有很好的规划推理能力，通过自上而下任务逐层分解，模块工作范围逐层缩小，问题求解精度逐层增高，实现了从抽象到具体、从定性到定量、从人工智能推理方法发展到数值算法的过渡，较好地解决了智能和控制精度的关系，其缺点是系统可靠性、鲁棒性、反应性差。

### 2. 包容结构

1986 年，布鲁克斯以移动机器人为背景提出了一种依据行为来划分层次和构造模块的思想[104]。他相信机器人行为的复杂性反映了其所处环境的复杂性，而非机器人内部结构的复杂

性,于是提出了包容结构(见图 15.2),这是一种典型的反应式结构(也称为基于行为或基于情境的结构)。包容结构中每个控制层直接基于传感器的输入进行决策,在其内部不维护外界环境模型,可以在完全陌生的环境中进行操作。Brooks 采用包容结构构造了多种机器人,这些机器人确实显示出非常强的智能行为。随后涌现了一批基于包容思想的研究成果。

包容结构中没有环境模型,模块之间信息流的表示也很简单,反应性非常好,其灵活的反应行为体现了一定的智能特征。包容结构不存在中心控制,各层间的通信量极小,可扩充性好。多传感信息各层独自处理,增加了系统的鲁棒性,同时起到了稳定可靠的作用。但包容结构过分强调单元的独立、平行工作,缺少全局的指导和协调,虽然在局部行动上可显示出很灵活的反应能力和鲁棒性,但是对于长远的全局性的目标跟踪显得缺少主动性,目的性较差,而且人的经验、启发性知识难以加入。

### 3. 三层结构

纯粹的分层递阶结构缺少对陌生环境的反应能力,单一的包容结构缺乏必要的理性和学习能力。20 世纪 90 年代初,3 个不同的研究小组几乎同时独立地提出了极其相似的解决方案——三层结构。三层结构由反馈控制层、慎思规划层和连接二者的序列层构成(见图 15.3)。同期提出的 3 种三层结构的差别主要在序列层上:文献[91]中的序列层是 REX/GAPPS 系统,文献[253]中的序列层基于 RAPs 系统,文献[158]中的序列层基于包容结构。

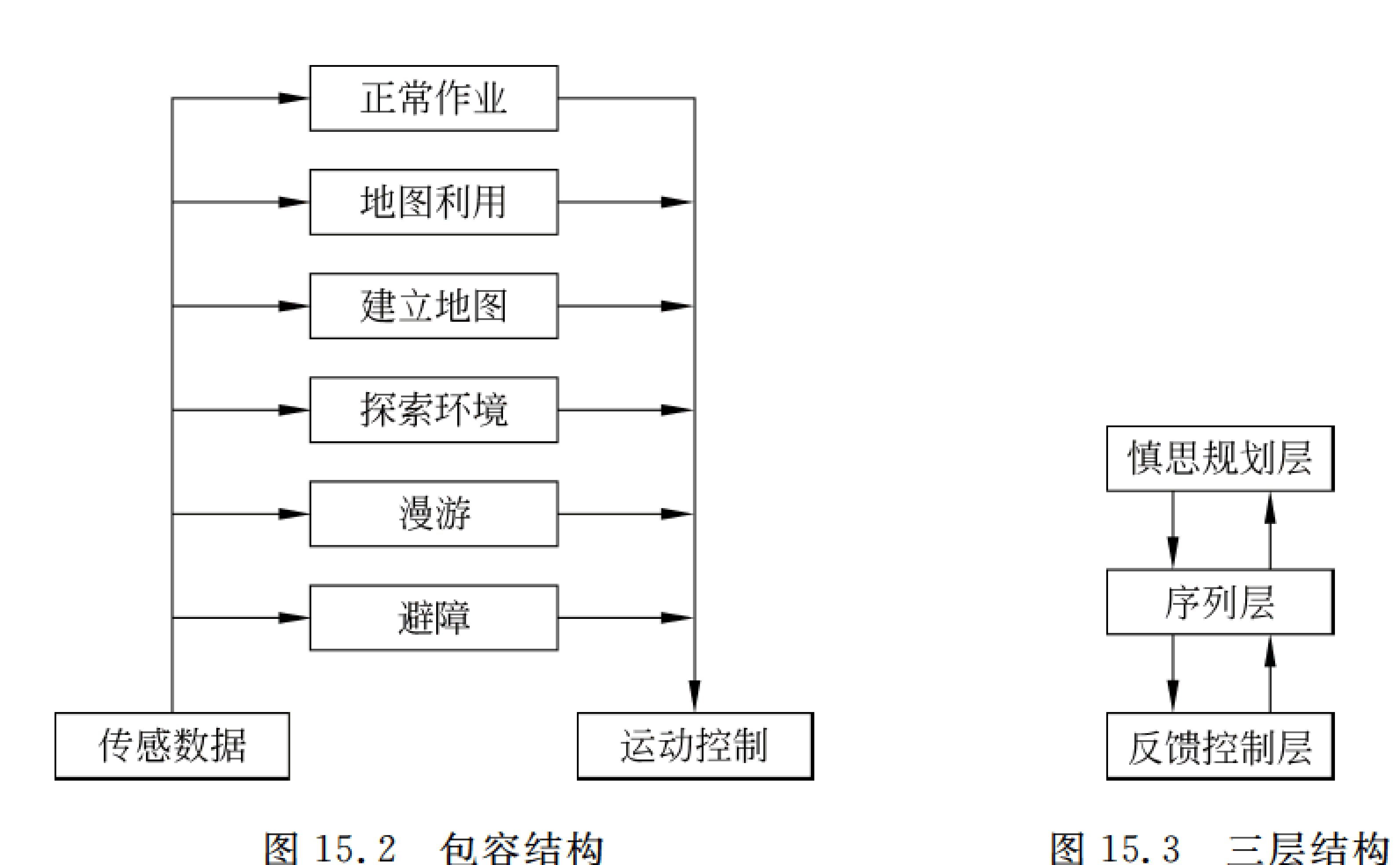



图 15.2 包容结构　　　　图 15.3 三层结构

三层结构是分层递阶和包容结构相融合的混合结构,它既吸取了递阶结构中高层规划的智能性,又保持了包容结构中低层反应的灵活性。机器人内部状态是传感信息融合的结果,是对外界环境的反映。三层结构中,序列层维护着状态信息,反映的是环境的过去,控制层直接处理传感信息,面对的是环境的现在,慎思层经过规划推理,预测的是环境的将来,从而保证了智能机器人在时间维上对环境的准确把握。三层结构的不足之处是忽视了传感信息融合、学习和环境建模(空间维)。以后几年实现的机器人采用层结构,都是基于三层结构进行改进或扩充的。

### 4. 自组织结构

1997 年,罗森勃拉特(J Rosenblatt)在移动机器人导航中提出了 DAMN 结构[633]。自组织结构如图 15.4 所示,它由一组分布式功能模块和一个集中命令仲裁器组成。各功能模块基于领域知识通过规划或反应方式自主产生行为(投票),由仲裁器产生一致的、理性的、目标导向的动作到控制器。各功能模块的投票受表决权大小的影响,表决权由模式管理器维护并可以动态修改。于是,在不同的任务、环境状态下,各功能模块会表现出不同的输入输出关系,即通过分布投票、集中仲裁且动态改变表决权的方式实现变构,从而使 DAMN 结构表现出自组织能力。

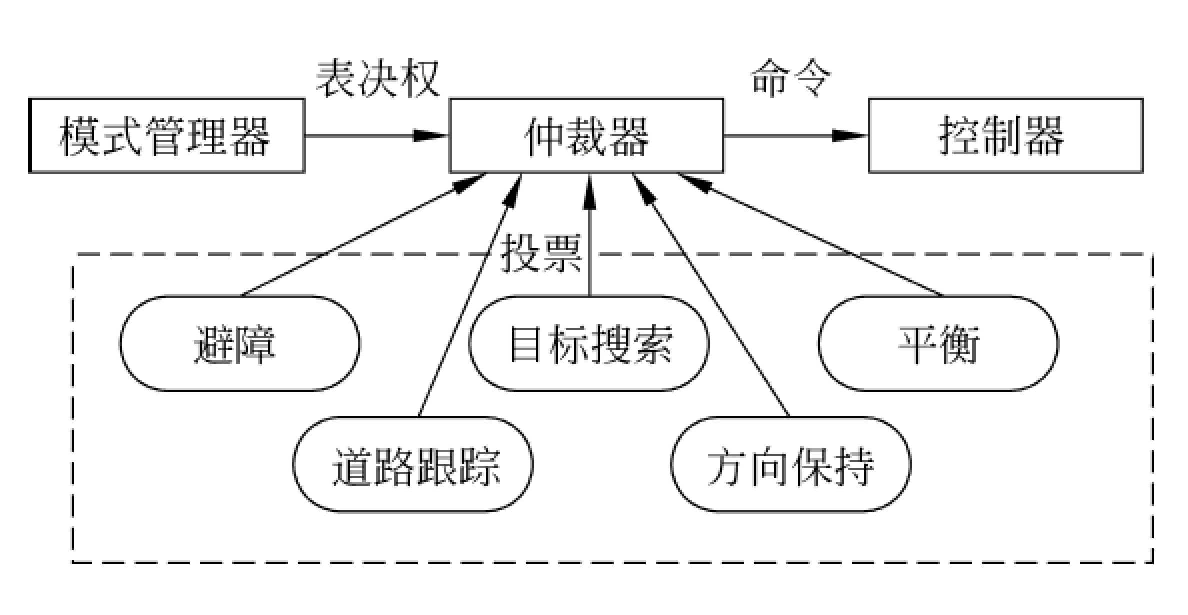



图 15.4 自组织结构

自组织结构的智能分布在其动态可变的结构中,突破了传统体系结构中功能分布模式固定的框架,具有良好的可扩充性和自适应、自组织性能,但其集中仲裁的机制往往是信息流通和系统控制的瓶颈。

### 5. 分布式结构

1998 年,比亚乔(M Piaggio)提出一种称为 HEIR(Hybrid Experts in Intelligent Robots)的非层次结构[574](图 15.5),由处理不同类型知识的 3 个部分组成:符号组件(S)、图解组件(D)和反应组件(R),每个组件又都是一个由多个具有特定认知功能的、可以并发执行的智能体构成的专家组,各组件没有层次高低之分,自主地、并发地工作,相互间通过信息交换进行协调,这是一种典型的分布式结构。

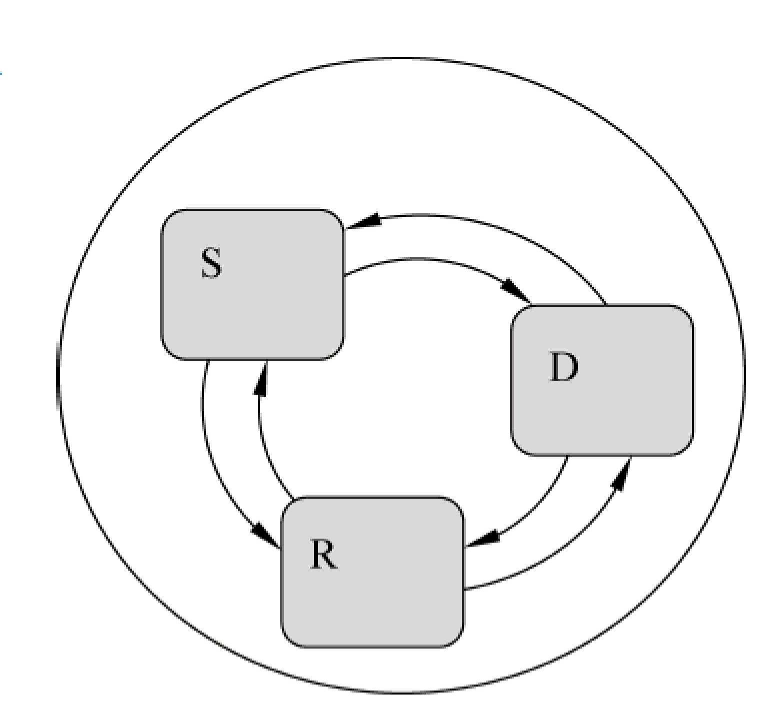



图 15.5 分布式结构

2000 年,盖苏(V Gesu)等在 DAISY 系统中实现了机器人自主导航的分布式体系结构[260],由 3 个独立的、可交互的智能体构成:目标识别器、障碍鉴别器和规划器。该结构是基于多智能体的,引入智能体交互与协作机制来处理机器人各模块之间的通信与协调。2002 年,科勒普(M Kolp)等将智能机器人控制系统划分为运动控制、传感、导航、环境建模和全局规划五个分布式组件,并分别以一个智能体来表示,采用组织理论中的五元结构来描述各智能体的分布模式、角色、任务、目标、资源、协调关系等[384]。

分布式结构突破了以往智能机器人体系结构中层次框架的分布模式,该结构中各个智能体具有极大的自主性和良好的交互性,可以独立求解局部问题并与系统中其他智能体通过交互保持协调,从而使机器人系统的智能、行为、信息和控制的分布具有极大的灵活性和并行性。但是,每个智能体对于要完成的任务拥有不全面的信息或能力,缺乏系统的和宏观的问题求解观念,难以保证智能体成员之间以及与系统的目标、意愿和行为的一致,对分散的共享数据和资源缺乏有效的分配和管理,冲突的检测和协调比较困难。分布式结构更多的适用于多机器人群体,机器人单体采用分布式结构,要建立必要的集中机制。

### 6. 进化控制结构

将进化计算理论与反馈控制理论相结合,形成了一个新的智能控制方法——进化控制。它能很好地解决移动机器人的学习与适应能力方面的问题。2000 年,蔡自兴提出了基于功能/行为集成的自主式移动机器人进化控制体系结构[849]。图 15.6 给出了进化控制体系结构,整个体系结构包括进化规划与基于行为的控制两大模块。这种综合的体系结构的优点是既具有基于行为的系统的实时性,又保持了基于功能的系统的目标可控性。同时该体系结构具有自学习功能,能够根据先验知识、历史经验、对当前环境情况的判断和自身的状况,调整自己的目标、行为,以及相应的协调机制,以达到适应环境、完成任务的目的。

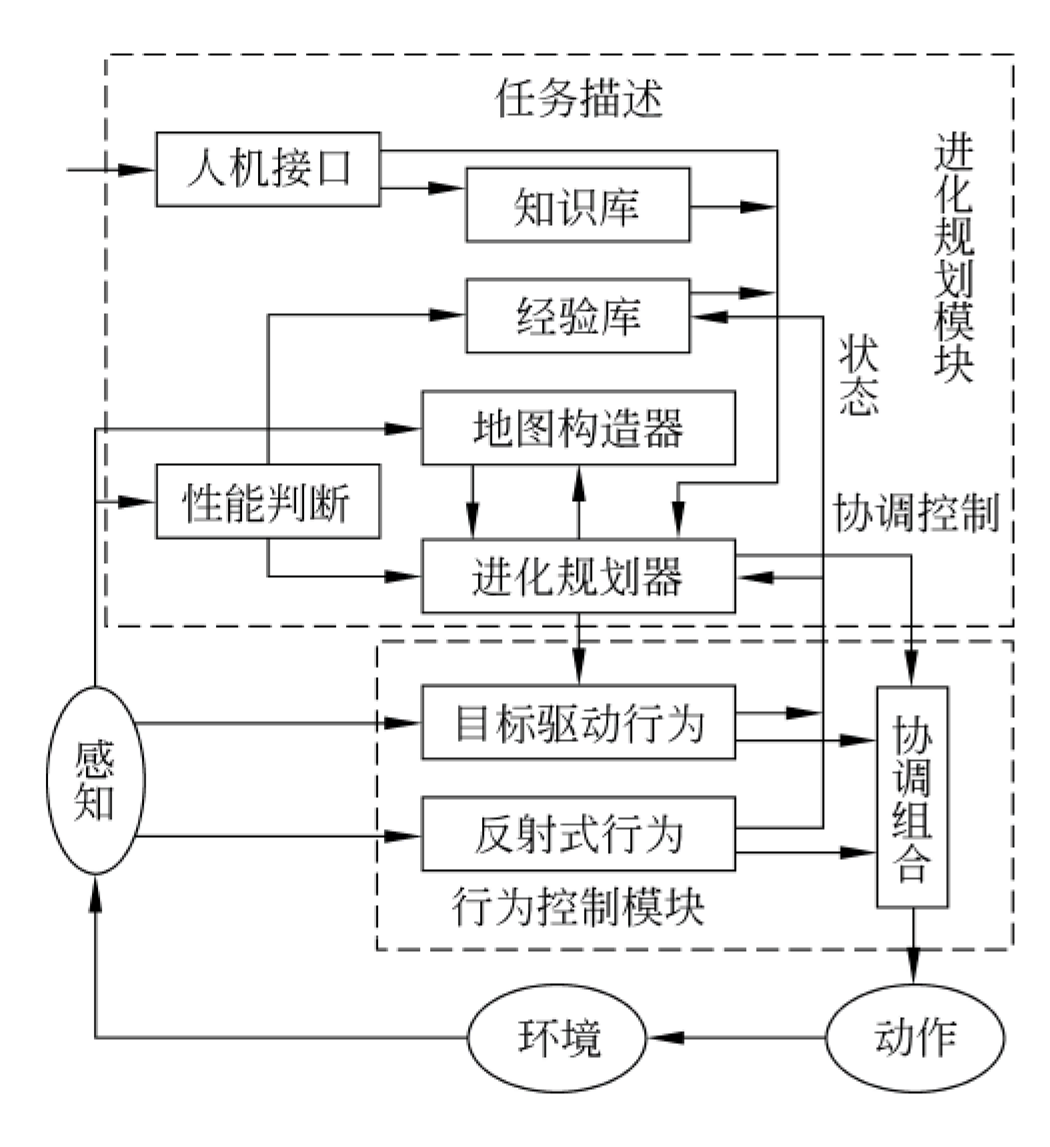



图 15.6　进化控制结构

进化控制结构的独特之处在于其智能分布在进化规划过程中。进化计算在求解复杂问题优化解时具有独到的优越性,它提供了使机器人在复杂的环境中寻找一种具有竞争力的优化结构和控制策略的方法,使移动机器人根据环境的特点和自身的目标自主地产生各种行为能力模块并调整模块间的约束关系,从而展现适应复杂环境的自主性。

### 7. 社会机器人结构

1999 年,鲁尼(B Rooney)等根据社会智能假说提出了由物理层、反应层、慎思层和社会

层构成的社会机器人体系结构[630]（图 15.7），其特色之处在于基于信念-愿望-意图（BDI）模型的慎思层和基于智能体通信语言 Teanga 的社会层，BDI 赋予了机器人心智状态，Teanga 赋予了机器人社会交互能力。

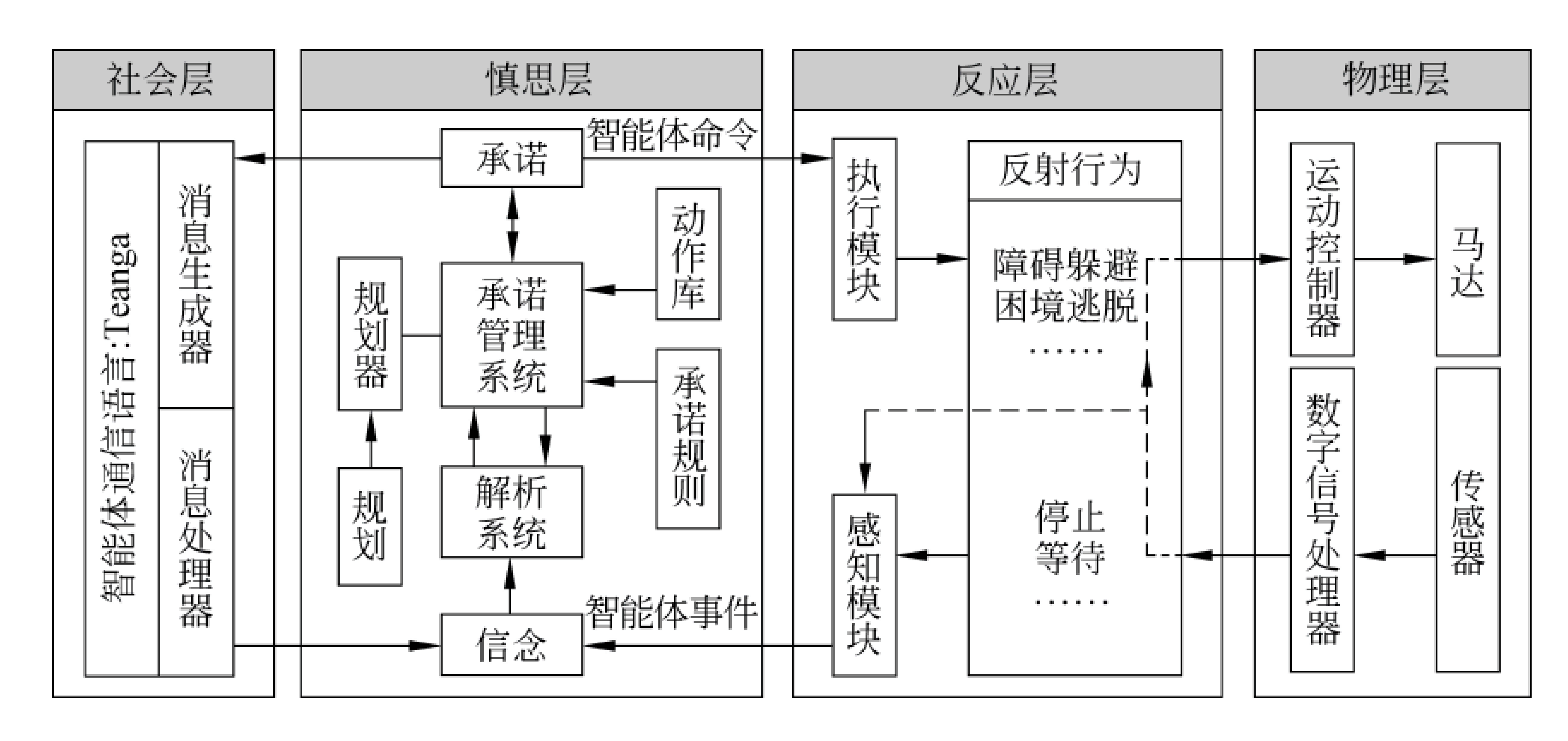



图 15.7　社会机器人的结构

社会机器人结构采用智能体对机器人建模，更自然，更贴切，能很好地描述智能机器人的智能、行为、信息、控制的时空分布模式，引入智能体理论可以对机器人的智能本质（心智）进行更细致地刻画，对机器人的社会特性进行更好地封装社会机器人结构继承了智能体的自主性、反应性、社会性、自发性、自适应性和规划、推理、学习能力等一系列良好的智能特性，对机器人内在的感性和理性、外在的交互性和协作性实现了物理上和逻辑上的统一。从人工智能到分布式人工智能，从智能体到多智能体，从单机器人到机器人群体，从人工生命到人工社会，智能科学正在经历着从个体智能到群体智能的发展过程。

### 8. 认知机器人结构

近年来，随着智能科学、行为学、生物学、心理学等理论成果的不断引入，认知机器人已成为智能机器人发展的一个重要课题。认知机器人是一种具有类似人类高层认知能力、并能适应复杂环境、完成复杂任务的新一代机器人。图 15.8 给出了一种认知机器人的抽象结构，分为三层，即计算层、构件层和硬件层。计算层包括知觉、认知、行动。知觉是在感觉的基础上产生的，是对感觉信息的整合与解释。认知包括行动选择、规划、学习、多机器人协同、团队工作等。行动是机器人控制系统的最基本单元，包括移动、导航、避障等，所有行为都可由它表现出来。行为是感知输入到行动模式的映射，行动模式用来完成该行为。在构件层包括感觉驱动器（感觉库）、行动驱动器（运动库）和通信接口。硬件层有传感器、激励器、通信设施等。当机器人在环境中运行时，通过传感器获取环境信息，根据当前的感知信息来搜索认知模型，如果存在相应的经验与之匹配，则直接根据经验来实现行动决策，如果不具有相关经验，则机器人利用知识库来进行推理。

密歇根大学的莱德（J Laird）等采用 SOAR 认知模型构建认知机器人[395]，系统中将符号处理与非符号处理结合，具有多种学习机制。多伦多大学的莱维斯克（H Levesque）等利用逻辑研究动态和不完全环境中认知机器人的知识表示和推理问题[428]。阿拉米（R Alarmi）等提出具有人类自我意识的机器人[9]。

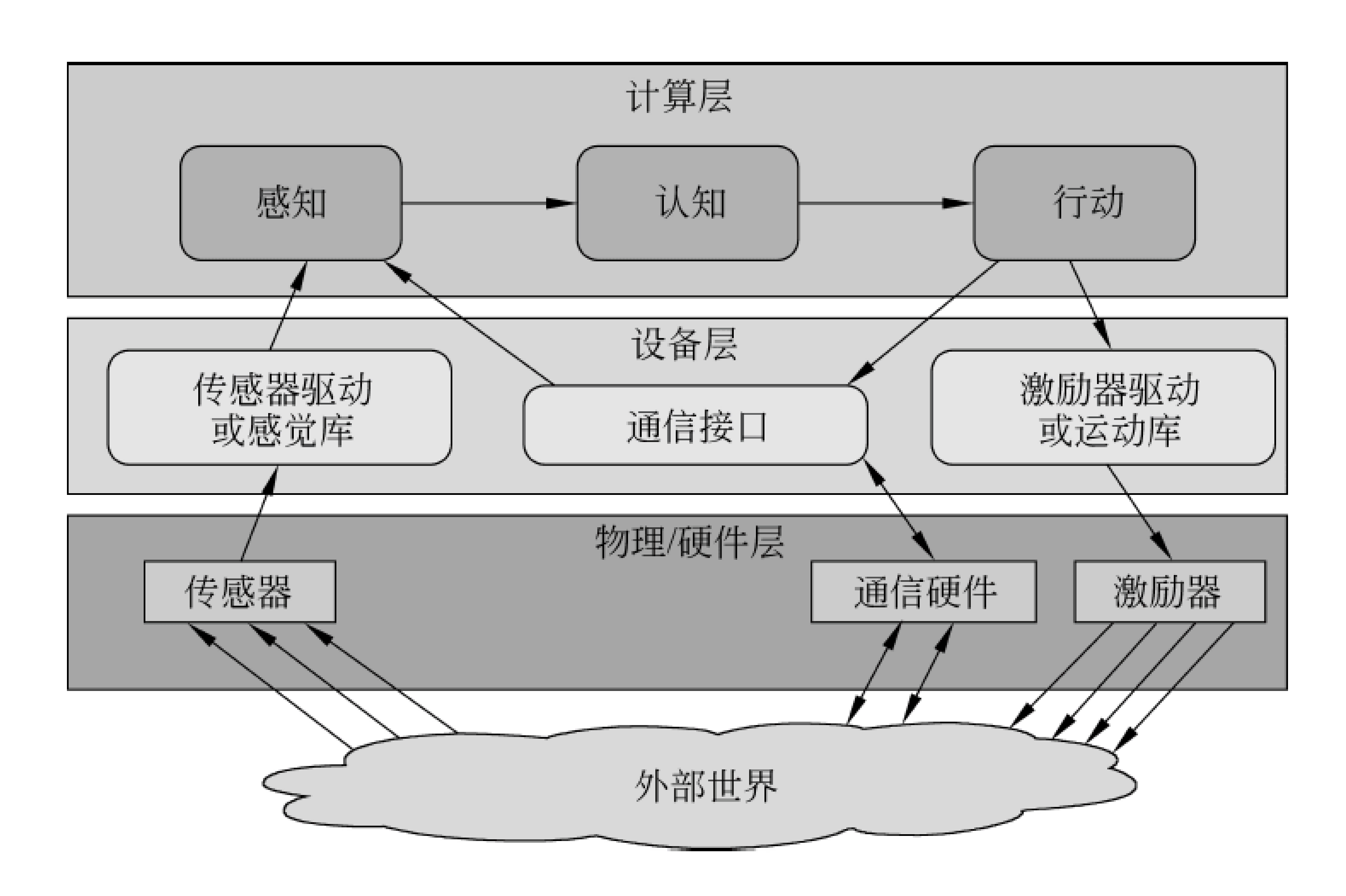



图 15.8　认知机器人的抽象结构

## 15.3　机器人视觉系统

机器人视觉系统是指用计算机来实现人的视觉功能,也就是用计算机来实现对客观的三维世界的识别。人类接收的信息70%以上来自视觉,人类视觉为人类提供了关于周围环境最详细可靠的信息。人类视觉所具有的强大功能和完美的信息处理方式引起了智能研究者的极大兴趣,人们希望以生物视觉为蓝本研究一个人工视觉系统用于机器人中,期望机器人拥有类似人类感受环境的能力。机器人要对外部世界的信息进行感知,就要依靠各种传感器。就像人类一样,在机器人的众多感知传感器中,视觉系统提供了大部分机器人所需的外部世界信息。因此视觉系统在机器人技术中具有重要的作用。

### 15.3.1　视觉系统分类

依据视觉传感器的数量和特性,目前主流的移动机器人视觉系统有单目视觉、双目立体视觉、多目视觉和全景视觉等。

#### 1. 单目视觉

单目视觉系统只使用一个视觉传感器。单目视觉系统在成像过程中由于从三维客观世界投影到 $N$ 维图像上,从而损失了深度信息,这是此类视觉系统的主要缺点。尽管如此,由于单目视觉系统结构简单、算法成熟且计算量较小,在自主移动机器人中已得到广泛应用,如用于目标跟踪、基于单目特征的室内定位导航等。同时,单目视觉是其他类型视觉系统的基础,如双目立体视觉、多目视觉等都是在单目视觉系统的基础上,通过附加其他手段和措施而实现的。

#### 2. 双目立体视觉

双目视觉系统由两个摄像机组成,利用三角测量原理获得场景的深度信息,并且可以重

建周围景物的三维形状和位置，类似人眼的体视功能，原理简单。双目视觉系统需要精确地知道两个摄像机之间的空间位置关系，而且场景环境的 3D 信息需要两个摄像机从不同角度，同时拍摄同一场景的两幅图像，并进行复杂的匹配，才能准确得到。立体视觉系统能够比较准确地恢复视觉场景的三维信息，在移动机器人定位导航、避障、地图构建等方面得到了广泛的应用。然而，立体视觉系统中的难点是对应点匹配的问题，该问题在很大程度上制约着立体视觉在机器人领域的应用前景。

### 3. 多目视觉

多目视觉系统采用三个或三个以上的摄像机，三目视觉系统居多，主要用来解决双目立体视觉系统中匹配多义性的问题，提高匹配精度。多目视觉系统最早由莫拉维克(H Moravec)研究，他为“Stanford Cart”研制的视觉导航系统[516]采用单个摄像机的“滑动立体视觉”实现；雅西达(M Yachida)[822]提出了三目立体视觉系统解决对应点匹配的问题，真正突破了双目立体视觉系统的局限，并指出以边界点作为匹配特征的三目视觉系统中，其三元匹配的准确率比较高；艾雅湜(N Ayache)[40]提出了用多边形近似后的边界线段作为特征的三目匹配算法，并用到移动机器人中，取得了较好的效果；三目视觉系统的优点是充分利用了第三个摄像机的信息，减少了错误匹配，解决了双目视觉系统匹配的多义性，提高了定位精度，但三目视觉系统要合理安置三个摄像机的相对位置，其结构配置比双目视觉系统更繁琐，而且匹配算法更复杂，需要消耗的时间更多，实时性更差。

### 4. 全景视觉系统

具有较大水平视场的多方向成像系统，其突出优点是具有较大的视场，可以达到 360°，是其他常规镜头无法比拟的。全景视觉系统可以通过图像拼接的方法或者通过折反射光学元件实现。图像拼接的方法使用单个或多个相机旋转，对场景进行大角度扫描，获取不同方向上连续的多帧图像，再用拼接技术得到全景图。美国南加州大学斯特恩(F Stein)[738]利用旋转摄像机获得 360°地平线信息为机器人提供定位信息；清华大学的刘亚[894]利用 360°旋转的摄像机拼接出镶嵌有运动目标的全景图，并对运动目标进行跟踪。图像拼接形成全景图的方法成像分辨率高，但拼接算法复杂，成像速度慢，实时性差。折反射全景视觉系统由 CCD 摄像机、折反射光学元件等组成，利用反射镜成像原理，可以观察周围 360°场景，成像速度快，能达到实时要求，具有十分重要的应用前景，可以应用在机器人导航中。日本大阪大学利用锥面反射镜研制出了 COPIS 全景视觉系统，为移动机器人提供定位、避障和导航。全景视觉系统本质上也是一种单目视觉系统，也无法直接得到场景的深度信息。其另一个缺点是获取的图像分辨率较低，并且图像存在很大的畸变，从而会影响图像处理的稳定性和精度。在进行图像处理时首先需要根据成像模型对畸变图像进行校正，这种校正过程不但会影响视觉系统的实时性，而且会造成信息的损失。另外，这种视觉系统对全景反射镜的加工精度要求很高，若双曲反射镜面的精度达不到要求，利用理想模型对图像校正则会存在较大偏差。目前，利用全景视觉最为成功的典型实例是 RoboCup 足球比赛机器人。

### 5. 混合视觉系统

混合视觉系统吸收各种视觉系统的优点，采用两种或两种以上的视觉系统组成复合视

觉系统,多采用单目或双目视觉系统,同时配备其他视觉系统。日本早稻田大学研制的机器人 BUGNOID 的混合视觉系统[192]由全景视觉系统和双目立体视觉系统组成,其中全景视觉系统提供大视角的环境信息,双目立体视觉系统配置成平行的方式,提供准确的距离信息;CMU 的流浪者机器人(Nomad)采用混合视觉系统,全景视觉系统由球面反射形成,提供大视角的地形信息,双目视觉系统和激光测距仪检测近距离的障碍物;清华大学的朱志刚[986]使用一个摄像机研制了多尺度视觉传感系统 POST,实现了双目注视、全方位环视和左右两侧的时空全景成像,为机器人提供导航。混合视觉系统具有全景视觉系统视场范围大的优点,同时又具备双目视觉系统精度高的长处,但是该类系统配置复杂,费用比较高。

## 15.3.2 定位技术

机器人研究的重点转向能在未知、复杂、动态环境中独立完成给定任务的自主式移动机器人的研究。自主移动机器人的主要特征是能够借助于自身的传感器系统实时感知和理解环境,并自主完成任务规划和动作控制,而视觉系统则是其实现环境感知的重要手段之一。典型的自主移动机器人视觉系统应用包括室内机器人自主定位导航、基于视觉信息的道路检测、基于视觉信息的障碍物检测与运动估计、移动机器人视觉伺服等。

移动机器人导航中,实现机器人自身的准确定位是一项最基本、最重要的功能。移动机器人常用的定位技术包括以下几个。

(1) 基于航迹推算的定位技术,航迹推算(Dead-Reckoning,DR)是一种使用最广泛的定位手段。该技术的关键是要能测量出移动机器人单位时间间隔走过的距离,以及在这段时间内移动机器人航向的变化。

(2) 基于信号灯的定位方法,该系统依赖一组安装在环境中已知的信号灯,在移动机器人上安装传感器,对信号灯进行观测。

(3) 基于地图的定位方法,该系统中机器人利用对环境的感知信息对现实世界进行建模,自动构建一个地图。

(4) 基于路标的定位方法,该系统中机器人利用传感器感知到的路标的位置来推测自己的位置。

(5) 基于视觉的定位方法。利用计算机视觉技术实现环境的感知和理解从而实现定位。

## 15.3.3 自主视觉导航

机器人自主视觉导航是目前世界范围内人工智能、机器人学、自动控制等学科领域内的研究热点。传统机器人自主导航依赖轮式里程计、惯性导航装置(IMU)、GPS 卫星定位系统等进行定位。而轮式里程计在车轮打滑情况下会产生较大误差,惯性导航装置(IMU)在长距离导航中受误差累积影响会造成定位精度下降,GPS 定位技术在外星球探测或室内封闭环境应用中受到诸多限制。因此,基于双目立体视觉的定位算法成为解决轮式里程计和惯性导航装置定位误差的可行方法。另外,机器人自主导航需要对周围环境进行实时动态的感知和重建,并构建地图用于导航和避障。传统的地形感知多使用激光雷达、声呐、超声、红外等传感器及相关方法,激光雷达功耗和体积较大,不适用于小型移动机器人,而超声、红

外传感器作用距离有限且易受干扰,但采用被动光学传感器的视觉方法,体积功耗小,信息量丰富,因此基于视觉方法进行地形感知与地图构建具有广阔的应用前景。

### 15.3.4 视觉伺服系统

最早基于视觉的机器人系统采用的是静态 Look and Move 形式。即先由视觉系统采集图像并进行相应处理,然后通过计算估计目标的位置来控制机器人运动。这种操作精度直接与视觉传感器、机械手及控制器的性能有关,这使得机器人很难跟踪运动物体。到 20 世纪 80 年代,计算机及图像处理硬件得到发展,使得视觉信息可用于连续反馈,于是人们提出了基于视觉的伺服控制形式。这种方式可以克服模型(包括机器人、视觉系统、环境)中存在的不确定性,提高视觉定位或跟踪的精度。

可以从不同的角度如反馈信息类型、控制结构和图像处理时间等方面对视觉伺服机器人控制系统进行分类。从反馈信息类型的角度,机器人视觉系统可分为基于位置的视觉控制和基于图像的视觉控制。前者的反馈偏差在 3D Cartesian 空间进行计算,后者的反馈偏差在 2D 图像平面空间进行计算。

从控制结构的角度,可分为开环控制系统和闭环控制系统。开环控制的视觉信息只用来确定运动前的目标位姿,系统不要求昂贵的实时硬件,但要求事先对摄像机和机器人进行精确标定。闭环控制的视觉信息用作反馈,这种情况下能抵抗摄像机与机器人的标定误差,但要求快速视觉处理硬件。根据视觉处理的时间可将系统分为静态和动态两类。

根据摄像机的安装位置可分为 Eye-in-Hand 安装方式和其他安装方式。前者在摄像机与机器人末端之间存在固定的位置关系,后者的摄像机则固定于工作区的某个位置。最近也有人把摄像机安装在机械手的腰部,即具有一个自由度的主动性。根据所用摄像机的数目可分为单目、双目和多目等。根据摄像机观测到的内容可分为 EOL 和 ECL 系统。EOL 系统中摄像机只能观察到目标物体;ECL 系统中摄像机同时可观察到目标物体和机械手末端,这种情况的摄像机一般固定于工作区,其优点是控制精度与摄像机和末端之间的标定误差无关,缺点是执行任务时,机械手会挡住摄像机视线。

根据是否用视觉信息直接控制关节角,可分为动态 Look-and-Move 系统和直接视觉伺服系统。前者的视觉信息为机器人关节控制器提供设定点输入,由内环的控制器控制机械手的运动;后者用视觉伺服控制器代替机器人控制器,直接控制机器人关节角。由于目前的视频部分采样速度不是很高,加上一般机器人都有现成的控制器,所以多数视觉控制系统都采用双环动态方式。此外,也可根据任务进行分类,如基于视觉的定位、跟踪或抓取等。

视觉伺服的性能依赖于控制回路中所用的图像特征。特征包括几何特征和非几何特征,机械手视觉伺服中常见的是采用几何特征。早期视觉伺服中用到的多是简单的局部几何特征,如点、线、圆圈、矩形、区域面积等以及它们的组合特征。其中点特征应用最多。局部特征虽然得到了广泛应用,而且在特征选取恰当的情况下可以实现精确定位,但当特征超出视域时则很难做出准确的操作,特别是对于真实世界中的物体。其形状、纹理、遮挡情况、噪声、光照条件等都会影响特征的可见性,所以单独利用局部特征会影响机器人可操作的任务范围,近来有人在视觉控制中利用全局的图像特征。如特征向量、几何矩、图像到直线上的投影、随机变换、描述子等,全局特征可以避免局部特征超出视域所带来的问题。也不需要在参考特征与观察特征之间进行匹配。适用范围较广,但定位精度比用局部特征低。总

之,特征的选取没有通用的方法,必须针对任务、环境、系统的软硬件性能,在时间、复杂性和系统的稳定性之间进行权衡。早期的视觉控制机器人,一般取图像特征的数目与机器人的自由度相同,例如威尔斯(Wells)和斯塔特森(Standersons)要求允许的机器人自由度数一定要等于特征数。这样可以保证图像雅可比是方阵,同时要求所选的特征是合适的,以保证图像雅可比非奇异。

## 15.4 机器人路径规划

移动机器人的路径规划是指机器人在有障碍物的工作环境中,寻找出一条从起点到终点的路径,使机器人在运动过程中能无碰撞地绕过所有障碍物到达目的地,其实质就是移动机器人运动过程中的导航和避障。基于不同的研究方向,移动机器人路径规划有着不同的划分标准。比较常用的有根据环境信息感知程度分类和根据环境信息确定性分类。根据环境信息的已知程度,可将路径规划划分为全局路径规划和局部路径规划以及两者相结合的情况。根据环境信息确定性程度,可以分为静态环境路径规划和动态环境路径规划。其中全局路径规划是在机器人工作环境信息已知的情况下离线规划出符合某种给定规则的最优路径,不需要考虑实时性问题。而局部路径规划中环境是未知的,可能存在动态障碍物。为了保证移动机器人的运行安全,不仅需要考虑机器人能够寻找到最优路径,还要考虑路径规划算法的实时性。移动机器人的主要工作环境在地面,在室内环境中容易受到灯光、走廊、门还有安装在地面的各种物品和工作人员移动的影响,而在室外环境中容易受到地面建筑物和行人的影响,因此对移动机器人在复杂环境下的路径规划问题仍是今后研究的重点之一。

在当今移动机器人路径规划中,全局路径规划主要是环境建模和路径搜索策略两个子问题。其中环境建模的主要方法有自由空间法、可视图法(V-graph)和栅格法(grids)等。路径搜索主要有 $A^*$ 算法、$D^*$ 最优算法等。局部路径规划的主要方法有遗传算法、人工势场法(Artificial Potential Field)、模糊逻辑算法和滚动窗口法等。

### 15.4.1 全局路径规划

全局路径规划算法(Global Path Planning)主要是指依据已获取的全局环境信息,给机器人规划出一条从起点至终点的运动路径。全局路径规划方法通常给出的是最优值,但是计算量大、时间久、实时性差,不适合动态环境下的路径规划,基于环境建模的全局路径规划算法主要有以下几种。

(1) 可视图法[620]。视机器人为一点,将机器人、目标点和障碍物的各顶点进行组合连接,要求机器人和障碍物各顶点之间、各障碍物顶点与顶点之间及障碍物各顶点和目标点之间的连线均不能穿越障碍物,即两点之间的直线是可视的。最优路径的搜索问题就转化为从起始点到目标点经过这些可视直线组合的最短距离问题。运用优化算法,删除一些不必要的连线以简化可视图,缩短搜索时间。该方法能够求得最短路径,但由于假设了机器人的尺寸大小忽略不计,使得机器人通过障碍物顶点时离障碍物太近,甚至发生接触并且搜索时间长,对于 $N$ 条连线的搜索时间为 TN。切线图法和 Voronoi 图法是对可视图法进行了改

进的方法。

(2) 自由空间法。在机器人路径规划的应用中,采用了预先定义形状的如广义锥形或者凸多边形等一些基本形状构造成为自由空间,并使用自由空间表示连通图,通过搜索连通图来进行路径规划[341]其优点是比较灵活,起始点和目标点改变不会使连通图出现重构的情况,其缺点是复杂程度与障碍物的多少成正比,在一些情况下无法获得最短路径。

(3) 栅格法。将机器人活动空间划分成一系列具有二值信息的网格单元,多采用四叉树或八叉树表示机器人活动范围[193]并通过优化算法完成路径搜索。该方法以栅格为单位记录环境信息,环境被量化成一系列具有一定分辨率的栅格,栅格的大小直接影响着环境信息存储量的大小和规划时间的长短。栅格划分大了,环境信息存储量小,规划时间短,但分辨率下降,在障碍物密集的环境下发现路径的能力减弱;栅格划分小了,环境分辨率高,在密集环境下发现路径的能力增强,但环境信息存储量迅速增大,且规划时间长。

## 15.4.2　局部路径规划

局部路径规划算法侧重于机器人探测的当前局部信息,这种机器人具有更好的实时性,其路径规划仅依靠传感器实时探测信息,现在很多机器人采用这种路径规划方法。这种路径规划有较强的实用性和实时性,对环境的适应能力强。其缺点是仅依靠局部信息,有时候会产生局部极小值或震荡,无法保证机器人顺利到达目标点。局部路径规划方法主要有以下几种。

(1) 遗传算法。由霍兰德(J Holland)于 20 世纪 60 年代初提出的[326],它以自然遗传、选择机制等生物进化理论为基础,利用选择、交叉和变异来培养控制机构的计算程序,在某种程度上对生物进化过程用数学方式进行模拟。它不要求适应度函数是连续可导的,而只要求适应度函数为正,同时作为一种并行算法,它的并行性适用于全局搜索。多数优化算法都是单点搜索算法,易于陷入局部最优值,而遗传算法是一种多点搜索算法,因而搜索到全局最优解的可能性更大。由于遗传算法的整体搜索策略和优化计算不依赖于梯度信息,所以很好地解决了其他一些优化算法无法解决的问题,但遗传算法运算速度不够快、进化元素众多,需要占据较大的存储空间和较长时间的运算。

(2) 人工势场法。它是由哈迪布(O Khatib)提出的一种虚拟力法,其基本思想是将机器人在环境中的运动视为一种虚拟的人工受力场中的运动[366]。障碍物对机器人产生斥力,目标点产生引力,引力和斥力的合力作为机器人运动的加速力,来控制机器人的运动方向和计算机器人运动速度。该方法结构简单,适宜于低层的实时控制,在实时避障和轨迹平滑的控制方面,取得了广泛的应用,但由于其存在局部最优解的问题,容易产生死锁现象,容易使使机器人在到达目标点之前就停留在局部最优点。

(3) 模糊方法。不需要建立完整的环境模型,也不需要进行复杂的计算和推理,尤其在对传感器信息的精度要求不高的情况下,机器人对周围环境和机器人的位姿信息具有不确定性、也不敏感,能使机器人的行为体现出很好的稳定性、一致性和连续性,能比较圆满地解决一部分规划问题。对处理未知环境下的规划问题显示出很大优越性,对于通常的用定量方法来解决很复杂的问题或当外界只能提供定性近似的、不确定信息数据时非常有效,但模糊规则往往是根据人们的经验预先制定的,所以存在着无法学习、灵活性差的缺点。

(4) 蚁群算法。它是由意大利学者杜里古(M Dorigo)等人于1991年创立的[195],是继神经网络、遗传算法、免疫算法之后启发式搜索算法的又一新发现。蚂蚁群体是一种社会性昆虫,它们有组织、分工,还有通信系统,它们相互协作,能完成寻找一条从蚁穴到食物源寻找最短路径的任务。人工蚁群算法是模拟蚂蚁群体智能的算法,具有分布计算、信息正反馈和启发式搜索的特点,在连续时间系统的优化和求解组合优化问题中获得广泛应用。

(5) 粒子群优化算法。这是一种进化计算技术,由埃伯哈特(Eberhart)和肯尼迪(Kennedy)于1995年提出的[363,364],他从鸟类捕食模型中得到启示并用于解决优化问题。在粒子群优化算法中,每个优化问题的解都是搜索空间中的一个粒子值,我们称之为"粒子",所有的粒子都有一个适应值是由被优化的函数决定的,每个粒子飞行的方向和距离都由一个速度决定,然后粒子们就追随当前的最优粒子在解空间中搜索粒子群优化算法,初始化为一群随机的粒子,在每一次迭代过程中,粒子通过跟踪两个"极值"不断更新自己。粒子群优化算法同遗传算法类似,是一种基于迭代的优化工具,但是并没有遗传算法用的交叉以及变异,而是粒子在解空间追随最优的粒子进行搜索。

(6) 滚动窗口。借鉴了预测控制滚动优化原理[966],把控制论中将优化和反馈两种基本机制合理地融为一体,使得整个控制既是基于反馈的,又是基于模型优化的。基于滚动窗口的路径规划算法的基本思路:在滚动的每一步,先进行场景预测,机器人根据其探测到的局部滚动窗口范围内的环境信息,用启发式方法生成局部子目标,并对窗口内动态障碍物的运动状态进行预测,判断机器人与动态障碍物相碰撞的可能性,机器人根据窗口内的环境信息及其预测的结果,选择路径规划算法,确定向子目标行进的局部路径,并依所规划的局部路径行进一步,窗口相应向前滚动,然后在新的滚动窗口产生后,根据传感器所获取的最新信息,对窗口范围内的环境信息及动态障碍物运动状况进行更新。该方法放弃了对全局最优路径的理想要求,根据机器人实时测得的实时局部环境信息,以滚动方式进行在线规划,具有良好的避碰能力,但存在着规划的路径是否最优路径的问题,也存在局部极小值问题。

## 15.5 细胞自动机

人工脑是在理解人脑的基础上,模拟和借鉴人脑的全部或部分结构和功能,采取进化的思想,研究开发的智能信息处理系统,解决传统计算机难以解决的复杂问题。它应具有拟人化、超高速、自进化、自组织、高度并行等特征。用计算机作为手段再现脑的思维决策过程,用人工脑控制器使机器人更聪明。人工脑通过学习使其能力不断进化,通过感知外界环境的刺激进行思考和决策,产生对外界环境的反应。人脑是生物长期自然进化的高度复杂而精巧的智能系统,建立具有类人智能的人工脑模型,还需要经过长期的艰辛努力和多学科专家的交叉研究。

从20世纪40年代以来,关于脑模型或人工脑的研究,人们已在仿生学、人工智能、人工神经网络、模式识别、超级计算机等领域进行了大量的探索,取得了一系列研究成果,其中感知机、联想机、细胞自动机、认知机等,都是某种简化的、局部的人工脑模型。

细胞自动机——仿脑机(CAM-Brain Machine,CBM)是1993年由日本京都先进电讯研究所(Advanced Telecommunications Research,ATR)的进化系统部(Evolutionary Systems Department)提出的人工脑计划,其目标是在8年内研制一个智商与小猫媲美的机器猫,包

含有 10 亿个神经元组成的人工脑。第一阶段人工脑的研制于 1999 年 3—4 月完工,并于 1999 年 11 月 9 日进行了正式展示。这一展示的机器猫的人工脑主要采用了人工神经网络技术,其中包含约 3770 万个人造神经细胞。尽管人工脑所含人工神经细胞数量与人脑的 1000 亿比相差甚远,但其智能超过昆虫[175]。

细胞自动机可视为由若干小单元构成的动态阵列,其中每一单元具有有限状态,在离散步序中,每一小单元按一致的法则,由其原状态及其邻域单元的状态决定新的状态。在任一时刻,诸小单元状态的总体构成细胞自动机的格局。从初始格局到最后格局的进化过程为计算处理过程,最后格局被视为计算结果。实际上,细胞自动机的每一小单元均为一有限态自动机,细胞自动机即为有限态自动机的动态阵列,细胞自动机格局的演化过程即为并行计算过程。细胞自动机的主要功能在于:可由局部特性及简单的一致性法则模拟、处理总体上具有高复杂性的离散过程和现象。

细胞自动机的小单元在细胞自动机结构中称为细胞。一个细胞自动机定义为有规律地分布于 $n$ 维空间中的一个细胞空间或细胞集合。$n=1$ 是最简单的情形,细胞分布在一条直线上,每一细胞为一方格,每一细胞有两个邻域细胞(简称邻域),每一细胞仅有两个不同状态,可由 0 和 1 值表示。在每一时刻,全部细胞状态值构成的序列为细胞自动机在该时刻的格局。在 $n=2$ 的情形,细胞在二维平面上分布,它们可为连续的方格、等边三角形、蜂房形式或其他形式。细胞空间为 $z^2$,其中 $z$ 为整数。每一细胞 $x=(x_1,x_2)$ 直接连接它的 8 个邻域。一般地,细胞自动机的基本模型具有 5 个主要特征。

(1) 它们由细胞的离散格局构成。

(2) 它们在离散时间步序中演化。

(3) 每一细胞的状态均在同一有限集中取值。

(4) 每一细胞的状态依同一确定的法则演化。

(5) 细胞状态的取值法则仅依赖于其自身及其周围局部领域细胞的状态值。

细胞自动机——仿脑机中采用收集发布模型(Collect or Disctribute Model,CoDi 模型)[257]。CoDi 模型中状态表示如图 15.9 所示。该模型中,一个单独的细胞具有 4 种基本类型,即胞体、树突、轴突和空胞。

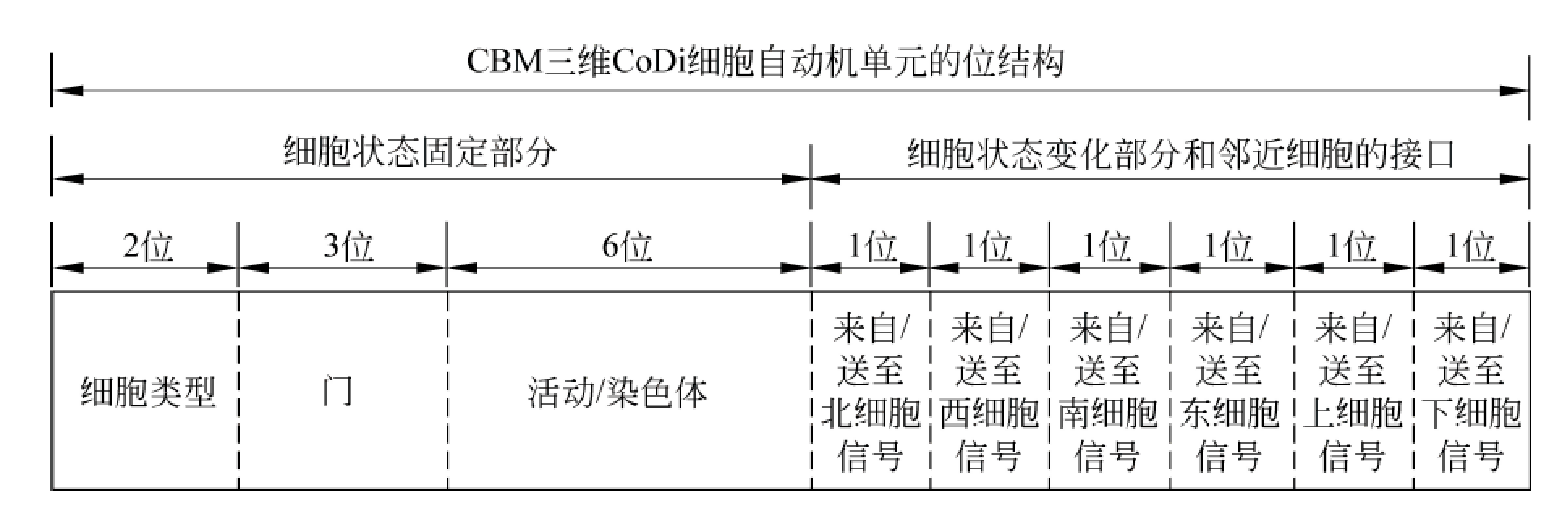



图 15.9 CoDi 模型中状态表示

(1) 空胞:在神经网络发信号时,不发生任何细胞的交互作用。

(2) 胞体:组成一个细胞自动机单元。胞体从周围轴突收集信号,按照内部规定的函数处理数据。在 CoDi 模型中神经元将信号求和,若超过阈值则点火。胞体的这种行为可以很方便修改以便符合给定的问题。胞体输出送到周围的树突。

(3) 树突:发布来自胞体的数据。

(4) 轴突:收集数据并送给胞体。

树突门指向接收神经信号的邻近细胞。树突仅接收来自那些邻近细胞的信号,但是它的输出可以提供给它的全部邻近细胞,通过这种方式树突发布信息。神经元是信息源。轴突接受邻近细胞的信息收集数据。它们给出的输出(如二进制的布尔或操作)仅到达门规定的邻近细胞。轴突通过这种方式收集和相加神经信号,直到收集的神经信号最终之和到达神经胞体。

每个树突和轴突都完全属于神经元胞体。图 15.10 给出了 CoDi 模型中神经信号处理的过程。每个神经元有两个轴突树和两个树突树。一个树突树发布抑制信号,另一个树突树发布兴奋信号。神经元有两个轴突树,在每个时间步它们最大限度地将信息送到神经元胞体。

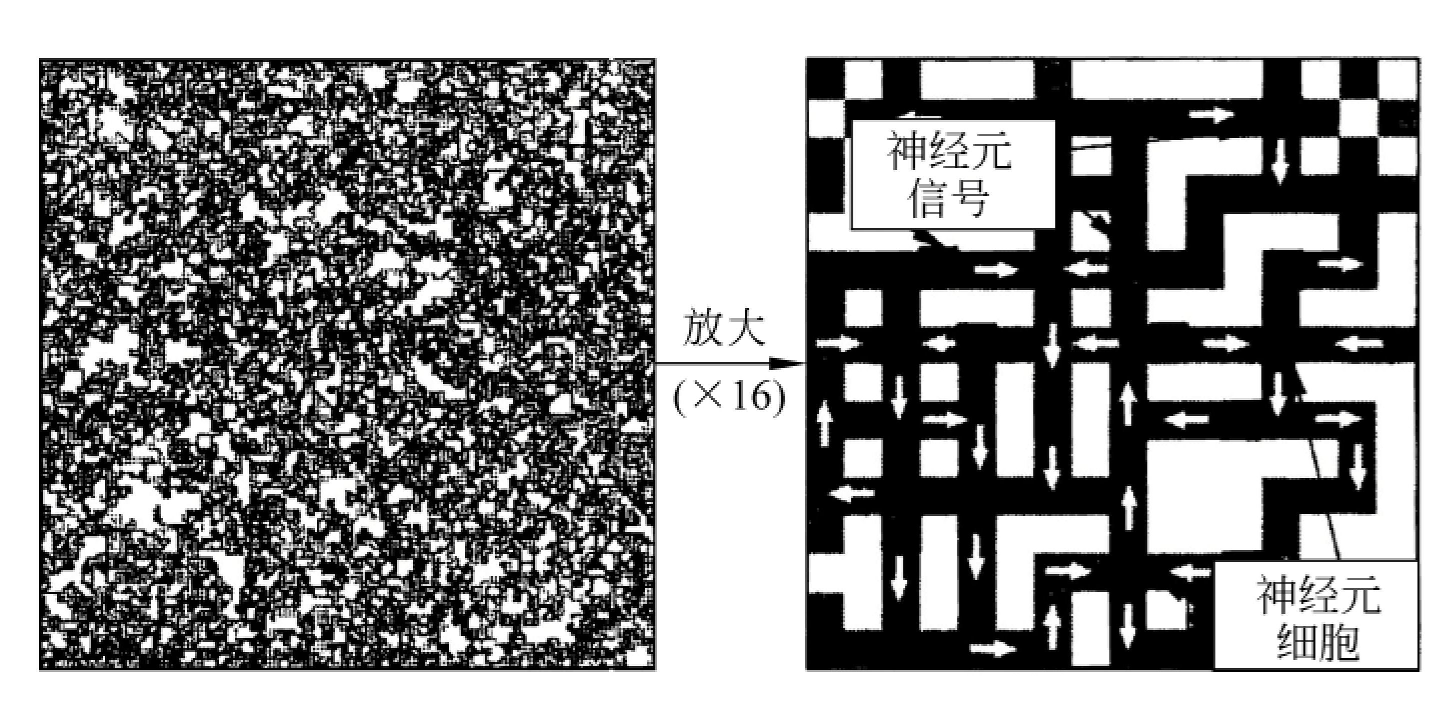



图 15.10 CoDi 模型中神经信号的处理

细胞自动机——仿脑机演示系统中细胞自动机由 72 块现场可编程门阵列芯片构成,每个芯片中包含大量晶体管,这些晶体管以几个为一单位组成所谓的细胞来模拟神经细胞,以及联系神经细胞的神经轴突和树突[176]。为了能执行一些特定任务,计算机神经网络必须具有学习能力,主要通过模拟自然大脑的生物进化过程来达到这一目的。细胞自动机中的现场可编程门阵列芯片具有与一般芯片不同的特性,它可不断重新设置内部电路,从而高效地对各晶体管之间连接关系进行反复改变和筛选。这些连接关系有点类似生物体中的染色体,当某种连接关系在执行特定任务为最佳选择时,这一染色体信息可遗传给下一代。经过代代选择,人工脑将自动获得最佳配置。

细胞自动机由以下 5 个子系统组成。

(1) 细胞自动机:工作空间采用两个相同的 CoDi 模块,每个模块有 13824 个细胞。

(2) 基因型/表现型记忆器:用于记忆染色体或神经网络描述。

(3) 适应度评估单元:用于评价神经网络。

(4) 遗传算法单元:用于执行遗传算法。

(5) 模块内联存储器:用于记忆网络中 64640 个 CoDi 模块间的连接[177]。

日本京都先进电讯研究所对于人工脑的研制正从硬件和软件两方面来推进。组合多个 CBM 的网络,若到 1000 个单元,则如由 10 亿个神经元组成的大脑一样工作,这也许能达到一只成年猫的智力水平。但这种方式并不正确。硬件进化是一个新的思路,即电子设备中元件之间的物理连接可以按照某种遗传算法进行自我更新。硬件的进化是一个极具挑战性的课题,而进化的硬件在目前仍处于开发的初期阶段。

## 15.6　认知机模型

美国麻省理工学院媒体实验室为了实现机器与人之间能流利地现场对话，提出构建认知机，使感知、动作、学习嵌入在系统中。罗易(Deb Roy)提出一种跨通道早期词汇学习模型(Cross-channel Early Lexical Learning，CELL)，理解幼儿从多种感知流早期获取词的过程[634]。

词汇学习是搜索跨输入通道结构，因此称作跨通道。CELL 构造创建词汇项模型，词汇项包括语言单元的规范说明和相应的感知基本类别的规范说明。CELL 是词汇获取过程的早期阶段。图 15.11 给出了跨通道早期词汇学习模型示意图。

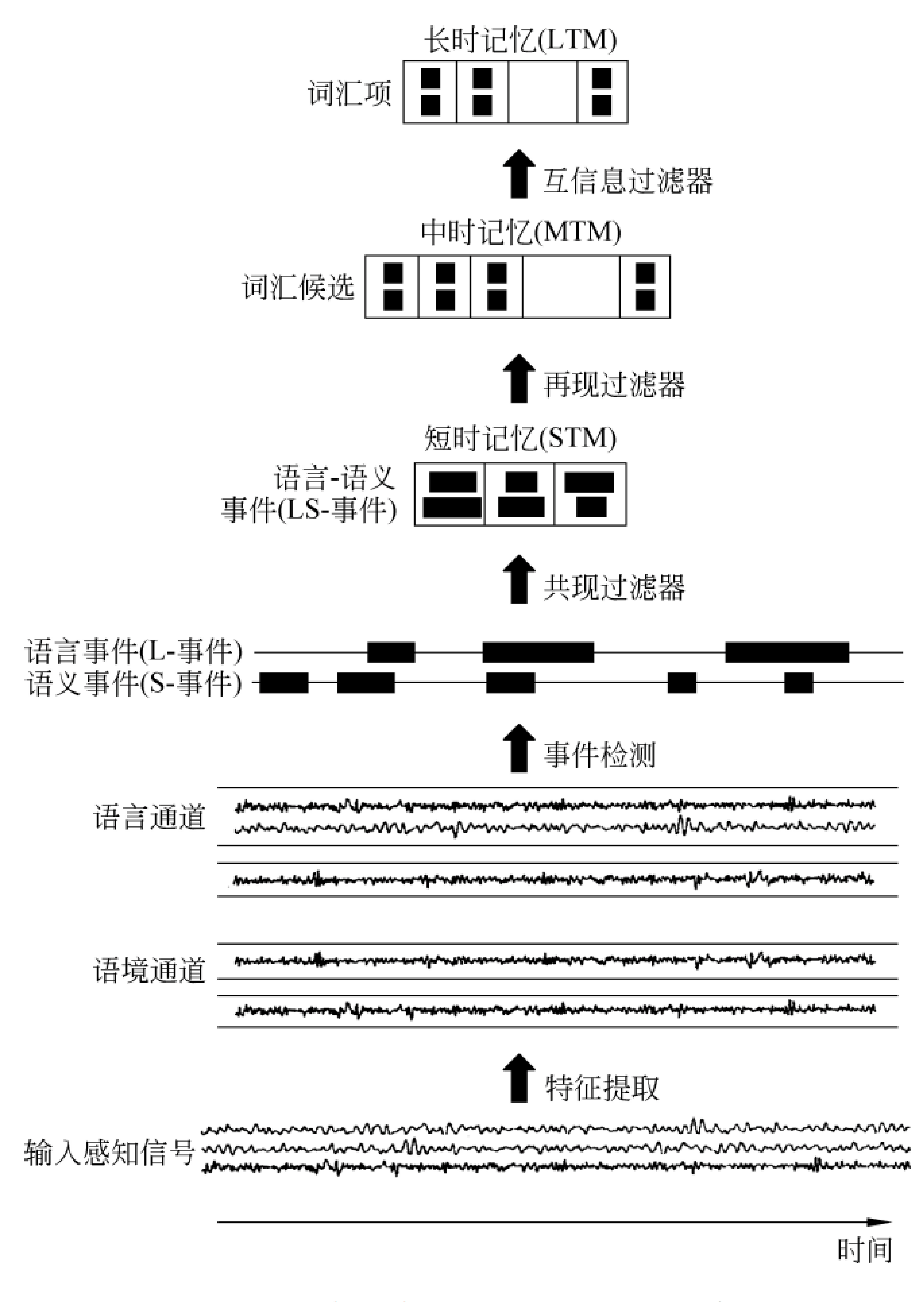



图 15.11　跨通道早期词汇学习模型示意图

语言通道包括：

(1) 按照音素或其他子词单元的话语声音。

(2) 话语音调轮廓。

(3) 可视嘴唇运动，帮助理解话语。

(4) 手势补充话语，对听觉不便的学习者提供基本的对话通道。

语境通道包括：

(1) 物体形状。

(2) 物体颜色。

(3) 物体大小。

(4) 物体之间的空间关系。

(5) 物体运动。

(6) 人脸标识。

在 CELL 模型中，跨输入通道检索实现下列操作：

(1) 假设与字对应语言单元原型。

(2) 假设语义类别的原型。

(3) 基于假设的原型最大化语言单元与语义类别的互信息。

(4) 基于导致高互信息的原型创建词汇项。

操作(1)和操作(2)将生成大量的可能的语言单元原型和相应的语义类别原型假设。操作(3)考虑每个原型对作为生成词汇项的基础。操作(4)选择最好假设作为基础，生成词汇项。

英国伯明翰大学的斯洛曼(Aaron Sloman)提出了一种认知影响机 CogAff 系统结构(见图 15.12)[727]。认知影响机 CogAff 分成三层：感知层、中央层、动作层。在感知层主要是反应机制，通过感知内外部条件，产生内外部相应的状态变化。动作层实现慎思机制，具有各种不同水平的抽象能力，灵活的、高效的记忆可用于思维、推测、部分规划、各种推理。目标生成处理、目标比较、规划构建、规划评价、规划执行的遇境依赖全局调整提供了基础，以便将来概念系列可以参考影响的状态和过程，这在纯反应系统结构中是不能实现的。第三层是元管理层，类似人的反射系统，利用类别和评价体系，对内部状态实现自观察、自控制，连接高级的学习和控制未来处理的机制。元管理层的操作如下。

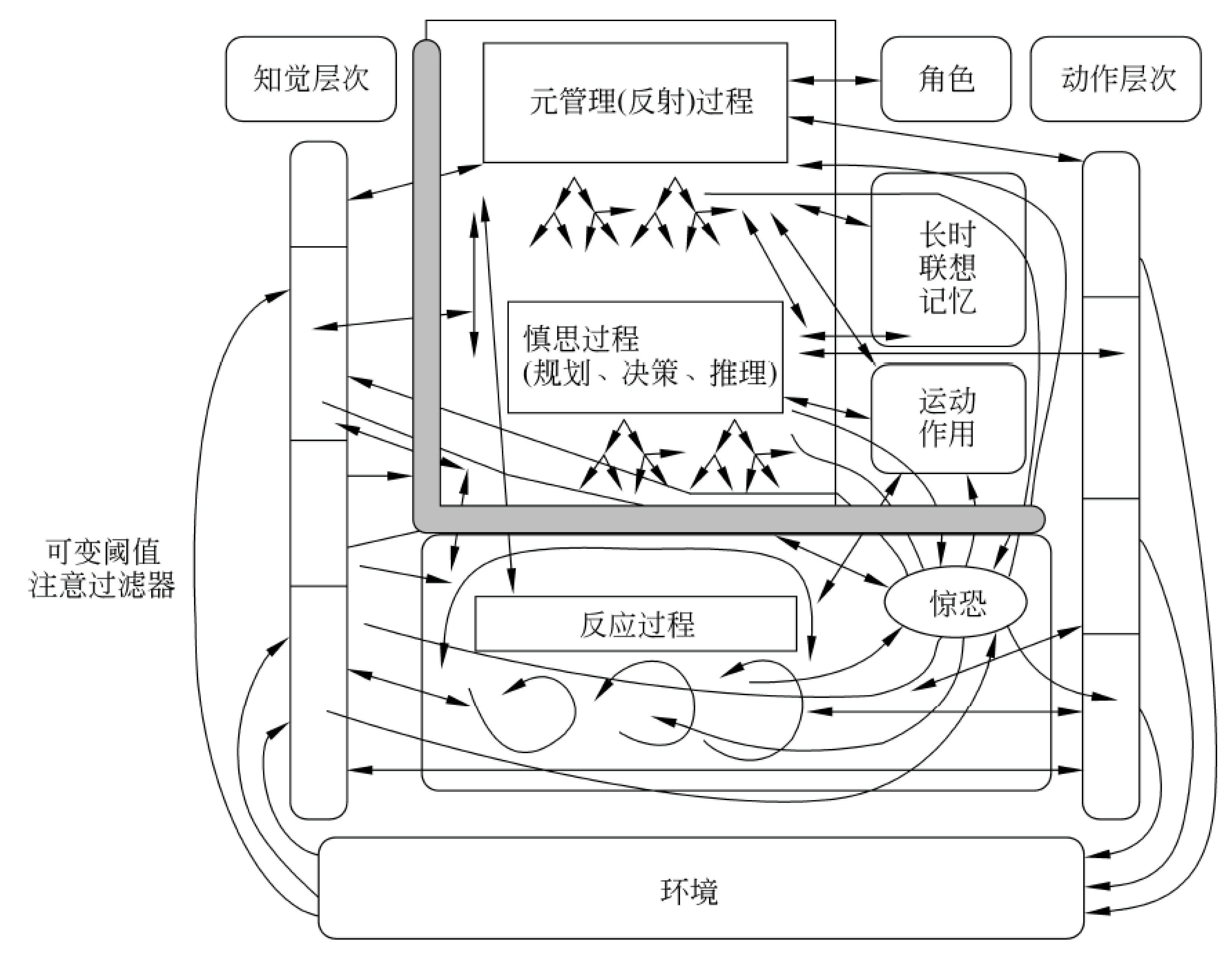



图 15.12 认知影响机 CogAff 系统结构

(1) 能够思考回答自己所想的和经历的问题。例如长方形和平行四边形。

(2) 能够通知或报告思想的循环过程。例如,为了达到 A 决定做 B,为了达到 B 决定做 C,为了达到 C 决定做 A,然后发现形成了一个圆圈思维。

(3) 能够根据事情的重要性判断自己所从事的事情。例如,晚上要完成家庭作业,而不看电视节目。

(4) 能够指导改变。例如,解决这个问题比前一个快,表明这次处理问题是多么正确。

## 15.7 情感机器人

情感机器人就是用人工的方法和技术赋予机器人以人类式的情感,使之具有表达、识别和理解喜乐哀怒,模仿、延伸和扩展人的情感的能力。

20 世纪 90 年代,各国纷纷提出了“情感计算”“感性工学”“人工情感”与“人工心理”等理论,为情感识别与表达型机器人的产生奠定了理论基础。主要的技术成果有:基于图像或视频的人脸表情识别技术;基于情景的情感手势、动作识别与理解技术;表情合成和情感表达方法和理论;情感手势、动作生成算法和模型;基于概率图模型的情感状态理解技术;情感测量和表示技术,情感交互设计和模型等。这种机器人能够比较逼真地模拟人的许多种情感表达方式,能够较为准确地识别几种基本的情感模式。

2008 年美国麻省理工学院开发出情感机器人 Nexi(见图 15.13),该机器人不仅能理解人的语言,还能够对不同语言做出相应的喜怒哀乐反应。Nexi 能够通过脸部自由活动装置与人沟通并通过面部表情表达高兴和忧郁之情。最令人感兴趣的是,这款机器人在表达情感过程中还能够通过转动和睁闭眼睛、皱眉、张嘴、打手势等形式表达其丰富的情感,可以根据人面部表情的变化来做出相应的反应。它的眼睛中装备有 CCD(电荷耦合器件)摄像机,这使得机器人在看到与它交流的人之后就会立即确定房间的亮度并观察与其交流者的表情变化。Nexi 还有灵敏的听觉系统。它的头臂都由高灵敏度的塑料制成,因此机器人能够感觉到别人触摸它,并根据当时情况对这种触摸是否带有恶意做出相应的反应。Nexi 可以做家务。

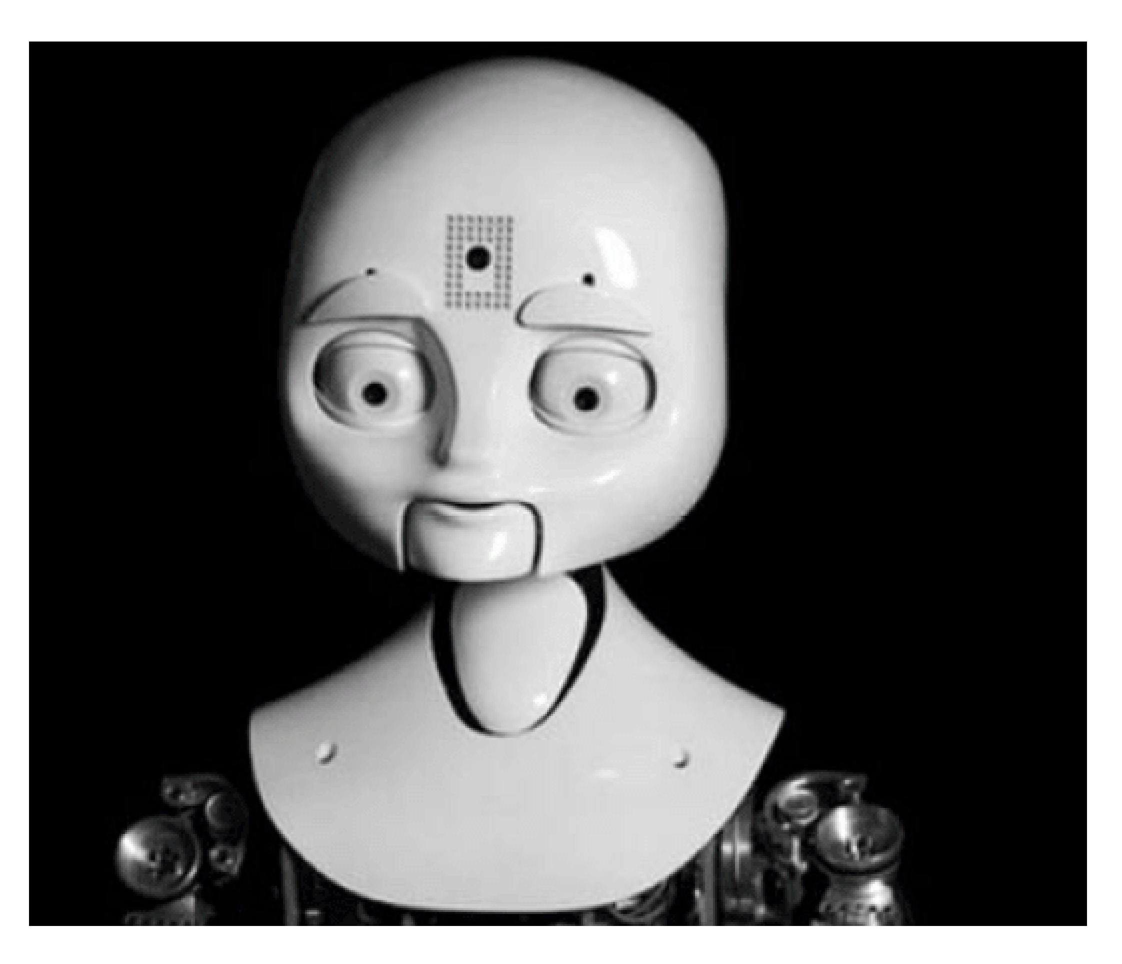


图 15.13 情感机器人 Nexi

韩国科学技术高等研究院智能机器人研究中心金中焕从另一种途径实现人工情感。他们为机器人编写出 14 个“人造染色体”(由单行的基因代码组成的软件程序),并把它植入到一条名叫 RITY 的虚拟机器狗身上。RITY 可以对外界环境产生带有情绪的反应,而且还可以做出与自己的“基因代码”所决定的“性格”相符的合理的决定,它可以对外界 47 种不同的刺激产生 77 种不同的行为反应。现在,金中焕正在试验如何把“人造染色体”放入硬件机器人中,让它们能够直接与人类进行情感交流。

欧盟第七研发框架计划(FP7)2010 年正式启动 ALIZ-E 具有情感的机器人研发项目,总投资 1060 万欧元。该课题由德国科学家领导,欧盟 5 个成员国德国、法国、意大利、荷兰和英国的科研人员组成研究团队,进行情感机器人研究,已完成实验室的预研工作,科研人员的研发工作以儿童和机器人之间的相互作用和相互学习为基础,研究模拟类似人类情感行为举止的情感机器人,正如人类情感源自对过去活动环境和经验积累的记忆互动以及相互接受。目前,情感机器人应用优化阶段的研发工作正在意大利米兰的 San Rafaele 医院小儿科进行。科研人员通过对儿童和机器人之间互动的跟踪观测,优化和调试机器人吸引和维持对儿童注意力的方式,从而自动重复产生机器人与儿童之间的共同行为举止、语言交流和游戏爱好等。科研人员希望应用优化阶段的情感机器人可以获得令人满意的结果,如此将在世界上首次制造出满足儿童需求的“伙伴机器人”。

真正具有类人情感的机器人必须具备三个基本系统:情感识别系统、情感计算系统和情感表达系统。

麻省理工学院的机器人专家布瑞兹(Cynthia Breazeal)创造了一个名为“克米特(Kismet)”的机器人,具有形状类似人头的情感系统(图 15.14)[545]。克米特配有可以活动的嘴唇、眼睛和眼睑,并且可以做出一系列的表情。当把它单独放在一边时,它看起来会显得十分忧愁,但是当感应到人类面部的时候,它就会做出笑的样子以引起人们的注意。如果有人推着它走得太快时,它甚至回流露出害怕的表情来提醒人们。和克米特玩耍时,人们都会不自觉地与它这些简单的感情流露产生共鸣。

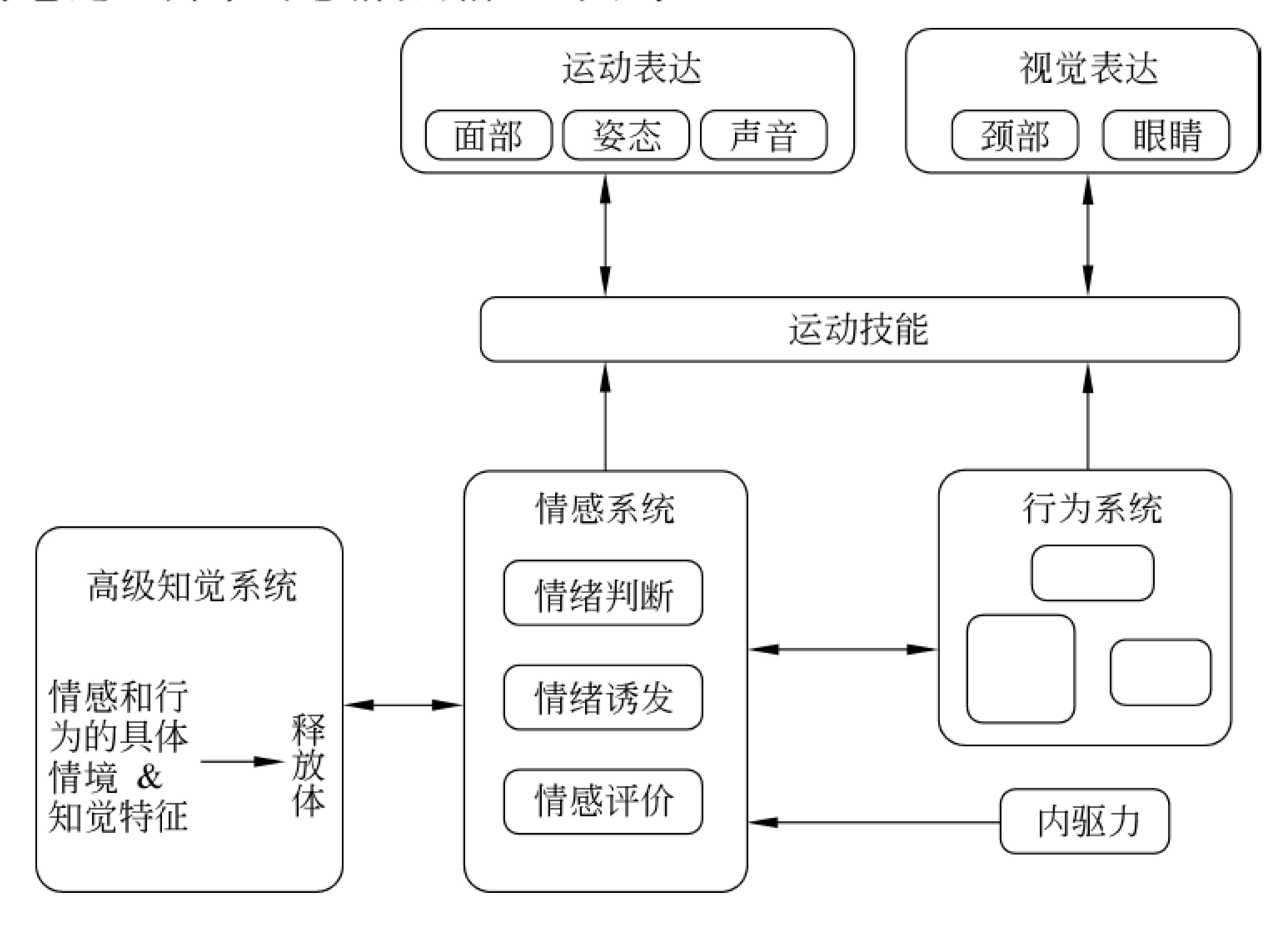



图 15.14 克米特的情感系统

尽管克米特是一个复杂的系统，具有身体和充当肌肉的多个发动机，以及基本的注意和情感模型，但它仍然缺乏真正的理解。因此，它向人们表现出的高兴和厌烦只是对环境中的变化的简单的程序反应，以及对动作和语音的物理特征的反应。当机器表达情感时，它们提供了与人交互的丰富的令人满意的活动，当然这种丰富和满意的解释和理解都来自人的头脑而不是人工系统。

## 15.8　发育机器人

1996 年，翁巨扬(J Weng)提出了机器人自主智力发育的思想[801]。2001 年，他在 Science 杂志上详细地阐述了自主智力发育的思想框架与可实现的算法模型[802]，即机器人在初始发育算法的控制下通过与环境的交流，动态地改变自己的记忆，对外界的刺激给出越来越积极的响应。

发育机器人与传统机器人的不同之处表现在：不是针对某种特定的任务，必须要对未知可能发生的任务生成合理的表示，要像动物一样可以在线的进行学习，同时这种学习是一种增量的过程，即要保证高层的决策建立在底层比较简单的基础之上。另外，自组织特性也是发育机器人的独特之处，在没有人类进行干扰的情况下，发育机器人也要保证能对所学知识进行合理的组织与存储。

发育模型的构建与发育学习算法的设计是发育机器人主要研究的两大方面。发育模型定义了从传感器信息获取到动作执行的一系列控制规则与算法，它包括以下四个部分(见图 15.15)。

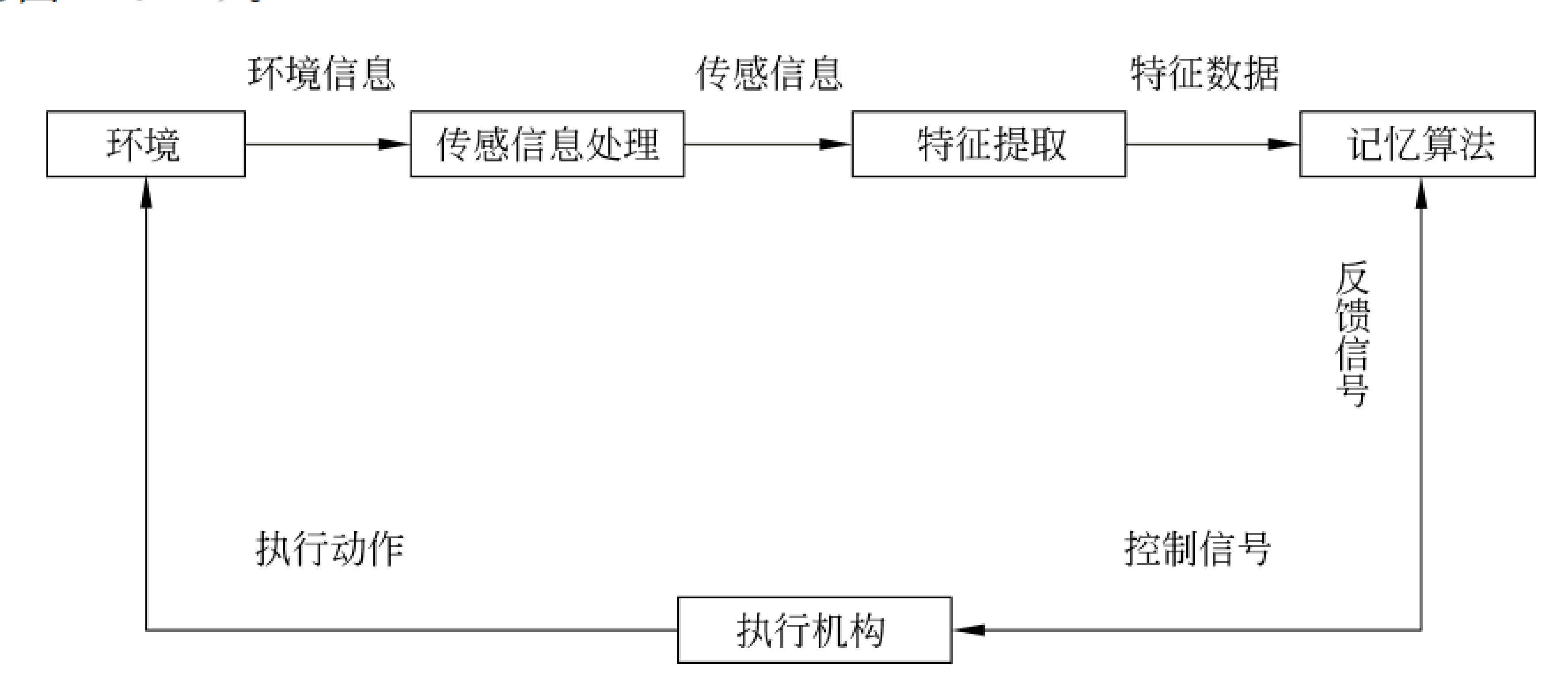



图 15.15　发育模型的基本结构

(1) 传感信息获取与预处理模块：传感器是机器人感知外界环境的窗口，只有装配了相应的传感器，机器人才能感知到相应的环境信息。因此，对传感信息进行处理是构成机器人智能的基础，发育机器人更是如此，因为机器人发育的过程就是其不断地与环境交互的过程。由于传感信息所含有的数据量非常巨大，且其中含有大量的噪声，所以对数据进行降维处理是非常必要的。

(2) 特征提取模块：特征提取算法既要保留原始数据的主要特征，又能将数据的存储量尽可能的大幅降低，是发育模型的一个必不可少的步骤。

(3) 记忆模块：记忆算法则是发育模型的核心所在，其相当于发育模型的中枢机构，因

为机器人在发育过程中所习得的经验均存储在这一结构之中。发育模型中的记忆算法要同时兼顾实时性与准确性的要求,同时要考虑到随着发育进程的深入,如何有效地降低存储量的问题。

(4) 执行模块:在记忆算法所输出的控制信号的控制下,对环境的变化做出反应,来完成各种不同的任务。

发育机器人模仿的是人脑及人心理发育的过程,需要机器人在实际的环境中自主的学习可用于完成各种任务的知识,并将这些知识有机地组织于记忆系统当中。因此,发育机器人研究者所面临的主要问题有:是否需要对环境建立具体的世界模型;能否对知识进行确定的表示;记忆系统如何组织以使记忆的提取能符合实时性的要求;机器人是否需要像生物一样,具有一些先天的条件反射机制;低层与高层的知识以何种方式进行组织,高层决策如何进行;多个传感器的数据如何进行融合(是否用到注意机制)以及采用何种学习方式等等。根据对以上问题回答的不同,研究者们提出了很多不同的发育模型,其中比较典型的有以下三种:CCIPCA+HDR 树模型、分层模型网以及模式(Schema)模型。下面分别介绍这几种模型。

### 1. CCIPCA+HDR 树模型

CCIPCA+HDR 树模型是由翁巨扬提出的[803],这种发育模型可以用于机器人的实时发育与自主增量学习。其主要包括两个基本的算法,即增量的主成分分析算法(CCIPCA)与分级回归树算法(HDR)。前者的输出作为后者的输入,可以实时地对环境的改变做出相应的反应。

主成分分析法(Principle Component Analysis,PCA)主要是对一系列输入的观察向量进行分析,找出最能表达这一向量组的少量正交基,实际上起到的就是对高维数据进行降维的作用,这样既可以保证不缺失原始特征,又可以有效地降低运算的复杂度,这对实时性要求较高的发育机器人来说尤为重要。但是一般的 PCA 方法需要对输入数据进行批处理,难以适应增量数据的要求,在这样的情况下,翁巨扬提出了增量的 PCA 方法即 CCIPCA 方法,它能够对依次输入的样本增量地计算主元,通过迭代的方法可以逐步收敛到待求的特征向量,其收敛性已从数学上得到证明。

HDR 算法则是一种针对高维向量子空间的识别与匹配算法。它采用了双重聚类的方法,可以自动区分输入样本,并根据其特征进行分类,将输入空间映射到输出空间,起到感知与动作匹配的作用。这种映射或者匹配对机器人而言,就是它们所学习到的知识。由于发育机器人实时的在环境中进行增量的学习,因此 HDR 树也是增量地建立的,随着 HDR 树规模的壮大,发育机器人也在不断地成长,具备更为细致的判别与区分的能力。

CCIPCA+HDR 树模型是基于判定树结构实现的,因此算法的时间复杂度为对数复杂度,满足了实时性的要求。同时与传统机器人相比,这一模型还具有较强的鲁棒性,可以适应有少量噪声的环境。该模型已经在密歇根州立大学的 SAIL 机器人平台上进行了导航、避碰、物体识别与语音识别等一系列实验,取得了较好的效果。但是这一模型缺乏高层决策与任务判别的能力,很难完成较为复杂的任务。另外,随着学习复杂程度的提高,存储量与计算量会大大增加,这对机器人的实时性与进一步发育都将会是一个不小的挑战。

## 2. 分层模型

布兰卡(D Blank)等提出了一种基于提取与预测机制的分层发育模型[81]。这种模型模仿了人类大脑皮层的工作机理,同时与 Brooks 的包容结构也非常相似。它将知识由低到高、由简单到复杂地组织在一个分层的结构当中,高层的知识建立于低层的知识之上。一些简单的底层控制由较低的层次来完成,这正如人类对熟悉的刺激所建立的条件反射一样,而对复杂烦琐的任务则要由高层的决策来实现。

分层发育模型的提取机制由自组织映射网络(SOM)来实现,而预测机制则采用简单的回归网络(SRN)。模型如图 15.16 所示,其中图 15.16(a)表示的是一个单层结构,从图中可以看出首先要对输入信号提取主要特征,随后根据这些主要特征来进行决策,而预测机制会根据上一步的决策对接下来的输入信号进行预测,预测准确率的高低代表了机器人对环境与任务的熟悉程度。图 15.16(b)将(a)中独立的结构组织在一个分层的模型当中,上层的提取模块会以下层模块提取的特征作为输入,每一层都能产生输出信号,这些信号被整合在一个包容结构当中,高层的决策优于低层的输出。这种模型既可以保证机器人对实时性的要求,又可以根据机器人的经历动态地改变其知识结构,体现了行为主义思想与发育思想的融合。

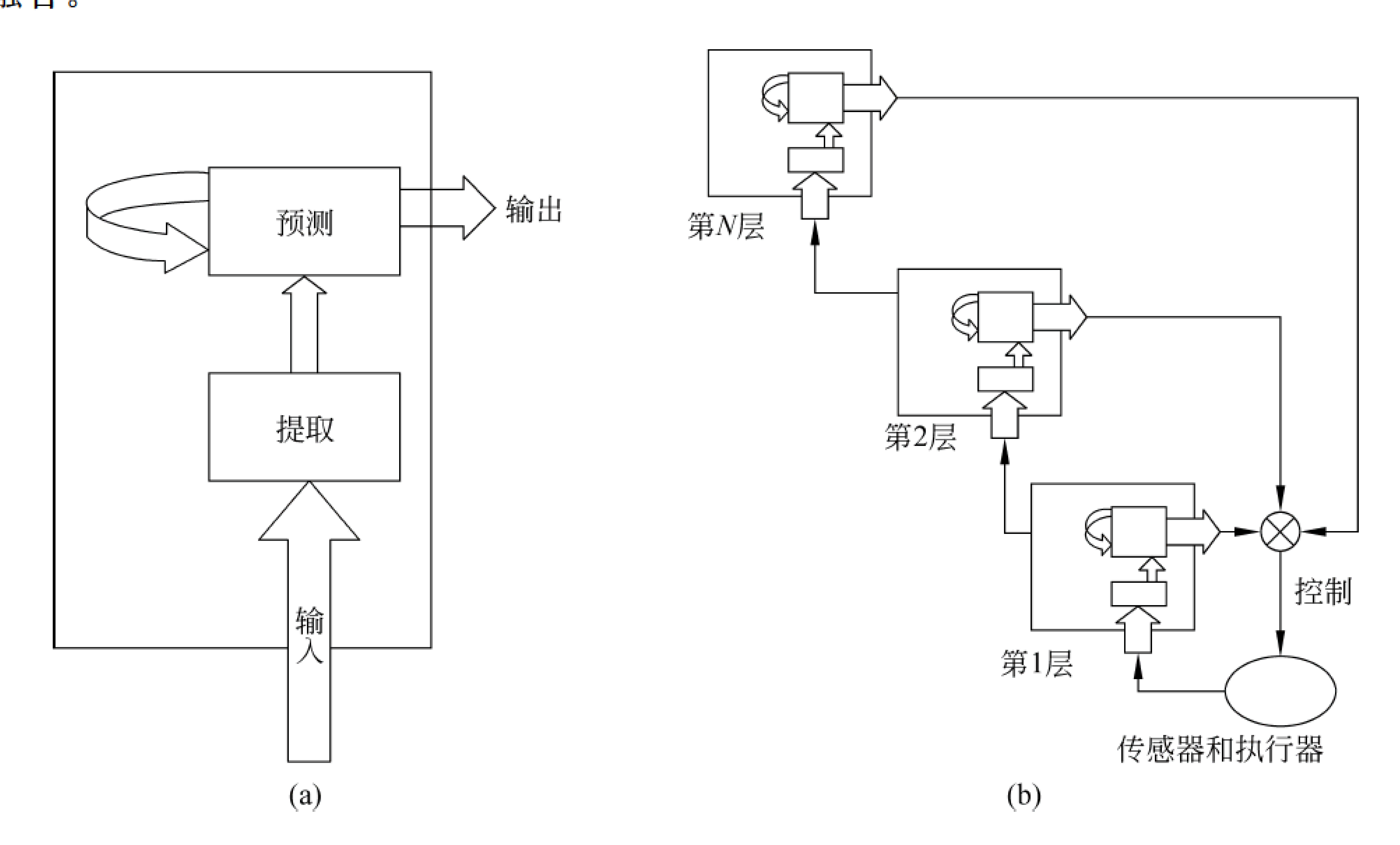



图 15.16 基于提取与预测机制的分层模型

分层模型具有良好的自适应性,从功能与结构上较好地模拟了人类的认知发育过程。这种模型存在的主要不足是结构复杂,高层决策的运算量过大,缺少对特定目标与任务的规划能力等等。

## 3. 模式模型

模式模型是由斯托雅诺夫(G Stojanov)提出的一种发育模型[740],其思想主要来源于发展心理学家皮亚杰的发生认识理论。发生认识论将人的认知发育划分为以下三个阶

段[568]：①通过遗传，具备先天的认知反应模式序列；②通过学习，可以修改原有的式序列，并生成新的可以更好适应环境的模式序列；③使自身逐渐适应这些新模式。

在模式模型中，首先要定义机器人的基本动作集 $A=\{a_1, a_2, \cdots, a_n\}$ 与基本感知集 $P=\{p_1, p_2, \cdots, p_j\}$，其中 $a_i$ 代表机器人所能采取的基本动作，而 $p_i$ 则代表机器人拥有的感知能力。随后要定义模式。模式实质上代表智能体有能力执行的一个基本的动作序列，如可以表示为 $s=a_1a_2a_5a_2a_7a_3$，它根据长度与动作种类的不同而有所区别。初始阶段，会自动生成基本的模式，在学习的过程当中，机器人试图执行这些基本的模式，但由于感知到环境的不同，相应的模式会进化为一个新的动作序列以适应环境与任务的需要。

模式模型已经在 Petitage 机器人上进行了导航方面的实验，取得了较好的效果。这个模型的特点是很好地模仿了人类认知的发育过程，具有较强的鲁棒性与自适应性，但是当感知的状态过多时，会极大地增加计算的时间复杂度，并影响算法收敛的速度。

## 15.9 智能机器人发展趋势

智能机器人从人类中成长，学习人们的技能，与人们拥有共同的价值标准，可以看成是人类思维的后代。新一代能力更强、用途更广的机器人称为“通用”机器人。莫拉维克曾在 20 世纪预测：第一代在 2010 年出现，它的明显特征是有多用途的感知能力以及较强的操作性和移动性；第二代在 2020 年出现，它最突出的优点是能在工作中学到技能，具有适应性的学习能力；第三代在 2030 年出现，这一代机器人具备预测的能力，在行动之前若预测到将出现比较糟的结果，它能及时改变意图；第四代会在 2040 年出现，这一代机器人将具备更完善的推理能力[515]。

根据 2016 年对 752 家机器人创业公司的报告显示，只有 25％的创业公司关注于工业机器人，而 75％的创业公司专注于解决新的机器人领域，如无人机、安保机器人、科研机器人、石油和天然气勘探机器人等，总共占 25％；农业领域机器人占 6％；AGV 占 7％；个人服务机器人占 3％；专业服务机器人占 7％；医疗、外科手术和康复机器人占 7％；消费产品机器人，如家庭清洁、安全、远程业务和娱乐占 9％；教育机器人占 5％。其他的企业还包括人工智能和软件、工程设计、元件制造、3D 打印技术、视觉系统和集成商。半数以上的初创公司主攻软件为基础的市场，而数据显示这些初创公司的硬件成本只占不到三分之一。未来 5～10 年智能机器人整体增长率都将超过两位数。智能机器人将重点发展如下几个方面。

### 1. 人机协作

汽车公司用人机协作机器人获得所需的灵活性，从而取代老式的工业机器人。优傲机器人和 Rethink Robotics 的公司网站包含了各种应用领域的人机协作机器人的使用案例视频。未来，这将成为这些新兴公司与库卡、ABB 等其他传统机器人巨头激烈竞争的机器人领域。这些新的人机协作机器人的主要特点体现在灵活性、安全性和易用性上。

在 2016 年 6 月举行的慕尼黑国际机器人及自动化技术贸易博览会上，每个机器人制造商都在展示旗下的人机协作机器人。把协作型机器人视为一种商品，可能并不会为公司获取更大的利润，但它创造的商业模式非常好。

### 2. 机器人即服务

许多初创公司正在寻找扩大经济规模效益的服务提供者。很多公司通过无人机捕捉传感器和相机数据，然后利用软件来分析数据，并将其转化为可操作的方案，这已经实现了跨界融合。另外，在物理空间真实操作的机器人和在软件中执行虚拟任务的机器人的界限也变得越来越模糊。因此，许多公司和服务提供商除了提供 SDK 之外，还开放其 API(应用协议接口)，这样这些新的机器人便可以扩大他们的使用范围和有效性，使它更容易为他们的用户服务。例如，比利时的一家创业公司 ZoraRobotics 便使用了亚马逊的 Echo/Alexa 系统和他们自己研发的软件，并将其应用到很多实体机器人上，以为健康和养老市场提供服务。

### 3. 更强大的算法

由智能机器人主动收集、交付信息及洞察结果的技术正在促进从人类生成信息资产到机器生成信息资产的转变。而这些资产包括新内容、分析与业务流程知识本体、知识产权。智能机器人将完善和推进被称为“算法业务”的新型业务模式。这是一种涉及大量互联、各类关系及动态洞察的经济形态，它基于以算法形式呈现的连接、大数据和新知识产权来支持行动。

智能机器人的崛起与其他发展趋势相辅相成，并必将与这些趋势共同颠覆我们的业务方式。新兴的算法业务即是其中最重要的趋势之一，它将带动能够产生新收入的新业务模式，借助算法充分利用大量与互联和关系有关的大数据的动态洞察结果。此类业务模式与智能机器人之间的关系非常密切，它将各种技术与智能机器人的服务结合在一起。

### 4. 推理决策

智能机器人将是比人类还优秀的推理者。与人类相比，它们的推理速度至少要快一百万倍，并且有百万倍的短期记忆力。推理是在计算中普遍存在的概念。它能模拟其他任何计算，大体上它可以模拟完成调节系统和全局建模器的功能，或者完成应用程序本身的任务。

在推理器中对控制系统的模拟，比在计算机上直接运行控制器程序要慢许多。但是，推理器能够抽象地检查模拟过程，设计出完成复杂操作的快捷步骤。通过不断地优化自身，推理器的控制器行动将变快，或许最终能比直接控制器还要快。它能加深对未来的预见，考虑意外事件的范围也会更广。

人类还有可能研制出完全基于推理的智能机器人。在这些机器人中，即使对于很小的行为也不是通过不灵活的调节反射进行的，而是朝着长期目标仔细规划实现的。冗长的意外事故链增大了有干扰事件出现时进行及时处理的可能性。

### 5. 依托云计算

人们正在探索云计算基础设施上的机器人，依靠云计算机处理大量的数据。这种方法可以直接调用“云机器人”，将使机器人不需要做复杂的计算，如图像处理和语音识别，甚至可以立即下载新的技能。很多相关项目都在进行中。特别是，谷歌有一个小团队，创建机器

人云服务,如果这种技术流行起来,对该领域可能是结构性的转变。在欧洲,一个重大项目 RoboEarth 的目标是开发“机器人万维网”,即一个巨大的云数据库,机器人可以共享有关的对象、环境和任务的信息。

通过深入了解用户,主动帮助个人以及相关人群工作更加高效,微软将“微软小娜”(Cortana)发展成为微软推动重塑生产力的跨平台工作界面。“微软小娜”智能体将逐渐取代诸如微软 Windows 之类的“操作系统”,成为控制、协调与促进用户互动和交流的媒介。到 2020 年,微软的发展战略将以“微软小娜”为中心,而非 Windows。微软正在通过融合、再定义与交付以智能体为媒介的用户体验而积极增强其云端办公系统技术套包,它将远胜于以前内部实施的 Exchange、SharePoint 以及个人+软件 Outlook 和 Excel。

# 第 16 章<<
# 类脑智能

通过脑科学、认知科学与人工智能领域的交叉合作，加强我国在智能科学这一交叉领域中的基础性、独创性研究，解决认知科学和信息科学发展中的重大基础理论问题，创新类脑智能前沿领域的研究。本章重点概述类脑智能的最新进展。

## 16.1 概述

21 世纪是智能革命的世纪。以智能科学为核心、生命科学为主导的高科技，将掀起一次新的高科技革命——智能技术革命。特别是智能技术、生物技术与纳米技术相结合，研制具有生物特征的智能机，将是 21 世纪高技术革命的突破口。

智能革命将开创人类后文明史。与能量革命实现能量的转换与利用不同，智能革命实现智能的转换与利用，即人把自己的智能赋予机器，智能机把人的智能转换为机器智能，并放大人的智能；人又把机器智能转换为人的智能，加以利用。如果说蒸汽机魔术般地创造了工业社会，那么智能机也一定能奇迹般地创造出智能社会。

图灵以人脑信息处理为原型，于 1936 年提出了伟大的图灵机思想[771]，奠定了现代计算机的理论基础。图灵曾试图"建造一个大脑"，第一个提出要把程序放进机器里，从而让单个机器能够发挥多种功能。

从 20 世纪 60 年代以来，冯·诺依曼体系结构是计算机体系结构的主流。现有计算机技术发展存在下列问题。

(1) 摩尔定律表明，未来 10～15 年内器件将达到物理微缩极限。

(2) 受限于总线的结构，在处理大型复杂问题上编程困难且能耗高。

(3) 在复杂多变实时动态分析及预测方面不具有优势。

(4) 不能很好地适应"数码宇宙"的信息处理需求。每天所产生的海量里，有 80%的数据是未经任何处理的原始数据，而绝大部分的原始数据半衰期只有 3 小时。

(5) 经过长期努力，计算机的运算速度达到千万亿次，但是智能水平仍很低下。

我们要向人脑学习，研究人脑信息处理的方法和算法，发展类脑智能成为当今迫切需求。目前，国际上非常重视对脑科学和智能科学的研究。2013 年 1 月 28 日，欧盟启动了旗舰"人类大脑计划(Human Brain Project)"，未来 10 年投入 10 亿欧元的研发经费[3]。目标是用超级计算机多段多层完全模拟人脑，帮助理解人脑功能。2013 年 4 月 2 日，美国前总统奥巴马宣布一项重大计划，将历时 10 年左右、总额 10 亿美元的研究计划"运用先进创新型神经技术的大脑研究(Brain Research through Advancing Innovative Neurotechnologies，BRAIN)"，目标是研究数十亿神经元的功能，探索人类感知、行为和意识，希望找出治疗阿

尔茨海默氏症(又叫老年痴呆症)等与大脑有关疾病的方法。

IBM 承诺出资 10 亿美元用于其认知计算平台 Watson 的商业化。高通量测序之父罗思伯格(J Rothberg)和耶鲁大学教授许田成立了新型生物科技公司,结合深度学习和生物医学技术研发新药和诊断仪器技术。

随着欧、美等国相继启动各种人脑计划,中国也将全面启动自己的脑科学计划。2015 年 3 月 9 日,百度董事长兼 CEO 李彦宏在全国政协会议上发言,建议设立国家层面的“中国大脑”计划,以智能人机交互、大数据分析预测、自动驾驶、智能医疗诊断、智能无人飞机、军事和民用机器人技术等为重要研究领域,支持企业搭建人工智能基础资源和公共服务平台,面向不同研究领域开放平台资源。

“中国脑计划”已经筹备了两三年时间,初步形成开展脑认知原理的基础、脑重大疾病、类脑智能的研究格局。类脑计算和人工智能研究是“中国脑计划”的重要组成部分,将以类脑智能研发与产业化为核心,从“湿”“软”“硬”和“大规模服务”这四个方向展开。具体包括:构建脑科学大数据和脑模拟平台,解析大脑认知和信息处理机制,即通常意义上的生物实验(湿);发展类脑人工智能核心算法,研发类脑智能软件系统,如深度学习算法就是一个特例(软);设计类脑芯片和类脑机器人,研发类脑智能硬件系统,从各种智能可穿戴设备到工业和服务机器人(硬);开展类脑技术在包括脑疾病在内的重症疾病的早期诊断、新药研发以及智能导航、智能专业芯片、公共安全、智慧城市、航空航天新技术、文化传播等领域的应用研究,推动新技术产业化(大规模服务)。“中国脑计划”已获国务院批示,并被列为“事关我国未来发展的重大科技项目”之一。类脑智能研究将借鉴脑的多尺度结构及其认知机制,提出并实现受脑信息处理机制启发的智能框架、算法与系统。

智能科学的研究表明,类脑智能是实现人类水平的人工智能的途径。类脑智能将基于神经形态工程,借鉴人脑信息处理方式,打破冯·诺依曼架构束缚,研究具有自主学习能力的超低功耗新型计算系统,适合实时处理非结构化信息,增强人类感知世界、适应世界、改造世界的智力活动能力。

人脑是世界上最复杂、最高级的智能系统,功能强、效率高、功耗低、普适性好。我们要向人脑学习,研究人脑信息处理的方法和算法。本章选取大数据、认知计算、人脑计划、神经形态机、脑机融合等内容,展示类脑智能研究的思路和进展,推动类脑智能的研究。

## 16.2 大数据智能

大数据本质上是人类社会数据积累从量变到质变的必然产物,是在信息高速公路基础上的进一步升级和深化,提升智能系统水平的重要途径,对人类社会的发展具有极其重大的影响和意义。

大数据是一个体量特别大,数据类别特别多的数据集,并且这样的数据集无法用传统软件工具对其内容进行抓取、管理和处理。大数据首先是指数据体量(Volumes)大,一般在 10TB 规模左右,但在实际应用中,很多企业用户把多个数据集放在一起,已经形成了 PB 级的数据量。其次是指数据类别(Variety)多,数据来自多种数据源,数据种类和格式日渐丰富,包括半结构化和非结构化数据。接着是数据处理速度(Velocity)快,在数据量非常庞大的情况下,也能够做到数据的实时处理。最后一个特点是指数据真实性(Veracity)高,企业

越发需要有效的信息之力以确保其真实性及安全性。大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

美国政府在 2012 年 3 月正式启动“大数据研究和发展”计划，该计划涉及美国国防部、美国国防部高级研究计划局、美国能源部、美国国家卫生研究院、美国国家科学基金、美国地质勘探局 6 个联邦政府部门，宣布将投资 2 亿多美元，用以大力推进大数据的收集、访问、组织和开发利用等相关技术的发展，进而大幅提高从海量复杂的数据中提炼信息和获取知识的能力与水平。该计划并不是单单依靠政府，而是与产业界、学术界以及非营利组织一起，共同充分利用大数据所创造的机会。这也是继 1993 年 9 月美国政府启动“信息高速公路”计划后，国家层面在信息领域的又一次发力。联合国也发布了《大数据促发展：挑战与机遇》的白皮书。全球范围内对大数据的关注达到了前所未有的热度，各类计划如雨后春笋般纷纷破土而出。

随着大数据、云计算、物联网、智能科学等技术广泛应用，人们通过搜索引擎等获取信息，寻找知识，构建知识图；人类的各种社会互动、沟通，社交网络和传感器也正在生成海量数据；商业自动化导致海量数据存储，但用于决策的有效信息又隐藏在数据中，如何从数据中发现知识，大数据挖掘技术应运而生，实现从数据到知识的飞跃。

## 16.3 认知计算

认知计算(Cognitive Computing)是指模仿人类大脑的计算系统，利用计算模型模仿人类思维过程，让计算机像人一样思维。认知计算涉及使用数据挖掘、模式识别和自然语言处理的自学习系统，以模仿人类大脑的工作方式。认知计算的目标是创建自动的信息技术系统能够解决问题，而无须人的援助。认知计算系统利用机器学习算法，通过挖掘反馈给它们的信息数据不断获取知识。该系统完善寻找模式和处理数据的方法，使它们成为有能力预见新的问题和建模可能的解决方案。

在 2005 年 7 月，IBM 公司和瑞士洛桑理工学院宣布开展蓝脑工程研究[453]，对理解大脑功能和机能失调取得进展，并且提供在精神健康和神经病里探索解决棘手的问题的方法。在 2006 年年末，蓝脑工程已经创建了大脑皮质功能柱的基本单元模型。2008 年 IBM 公司使用蓝色基因巨型计算机，模拟具有 5500 万神经元和 5000 亿个突触的老鼠大脑。IBM 公司从美国国防部先进研究项目局(Defense Advanced Research Projects Agency，DARPA)得到 490 万美元的资助，研制类脑计算机。IBM Almaden 研究中心和 IBM Wason 研究中心一起，斯坦福大学、威斯康辛-麦迪逊大学、康奈尔大学、哥伦比亚大学医学中心和加利福尼亚 Merced 大学都参加该项计划研究。

2007 年以来，从针对小鼠和大鼠脑皮质规模的早期工作开始，IBM 项目组的模拟在规模方面一直保持稳步增长。2009 年 5 月，在与劳伦斯伯克利(Lawrence Berkeley)国家实验室的合作中，IBM 项目组使用黎明蓝色基因(Dawn Blue Gene/P)超级计算机系统，获得了最新的研究结果(如图 16.1 所示)。该研究成果充分利用了超级计算机系统的存储能力，是具有价值的猫科一规模脑皮质模拟(大致相当于人脑规模的 4.5%)的里程碑[19]。这些模拟网络展示了神经元通过自组织形成可重现且具有锁时特性的非同步分组[20]。

脑认知研究推动伴随大数据的认知计算时代到来。2014 年 3 月 6 日，IBM 宣布将其大数据分析平台更名为 Watson Foundations，“Watson(沃森)”将作为“认知计算”的代名词成

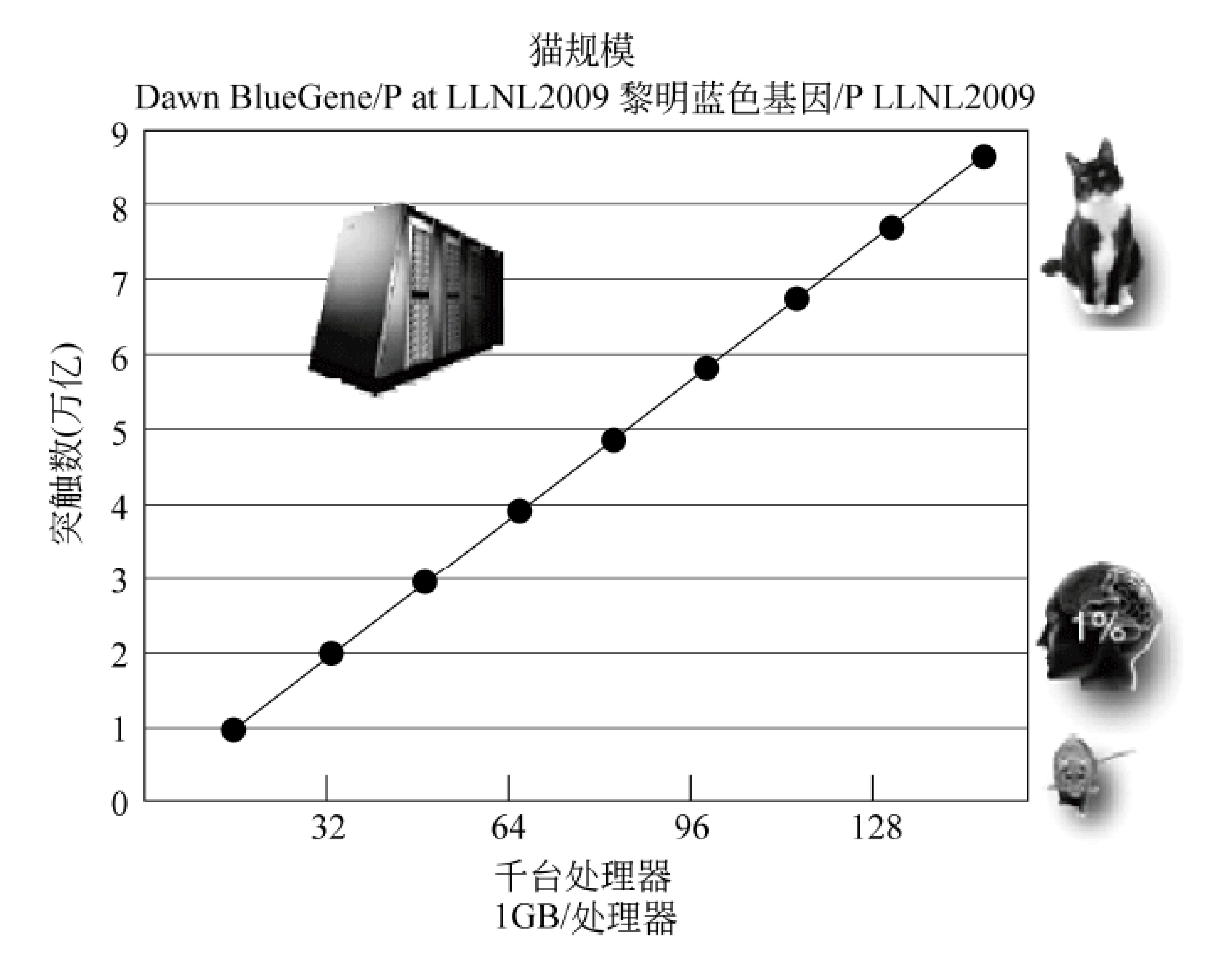



图 16.1 利用 C2 的可伸缩脑皮质模拟

为 IBM 未来的大数据战略方向。认知计算系统能够通过辅助(Assistance)、理解(Understanding)、决策(Decision)、洞察与发现(Discovery),帮助企业更快地发现新问题、新机遇和新价值。当你向沃森提问的时候,沃森处理的步骤如下。

(1) 将你的问题分解为很多关于这个问题的"特征"。

(2) 在大量可能包含答案的信息中进行搜索,然后生成一系列潜在答案。

(3) 使用特有的算法,为每一个潜在答案打分。

(4) 提供评分最高的答案,以及答案的相关证据。

(5) 对评分进行权衡,为每个答案的评分指数进行评估。

目前,IBM 把沃森系统应用到医疗行业。沃森能够对海量的医学数据进行处理和分析,快速了解患者情况,然后通过这些信息来进行"诊断",仅需几秒,就可以得出较为准确的结果,把问答集成到临床和业务决策中,为医生的最终诊断提供依据和帮助。

2015 年 1 月 29 日,在第 29 届美国人工智能大会上,IBM Watson 研究中心的塞尔曼(M Sellmann)作了"智能决策"的特邀报告[668]。报告指出人工智能技术的最新进展已经提供了商业上可行的协同战略决策支持系统愿景。这些认知系统集成信息检索、知识表示、交互式建模以及社会学习能力与逻辑推理在不确定条件下的概率决策。认知计算获得广泛应用,包括专家系统、自然语言编程、神经网络、机器人和虚拟现实。

## 16.4 欧盟人脑计划

2013 年 1 月 28 日,欧盟委员会宣布"未来和新兴技术(FET)旗舰项目"的竞选结果,人脑计划(Human Brain Project,HBP)将在今后 10 年中获得 10 亿欧元的科研资助。

人脑计划项目希望通过打造一个综合的基于信息通信技术的研究平台来研发出最详细的人脑模型。在瑞士洛桑联邦理工学院的马克拉姆(H Markram)的协调下,来自 23 个国家(其中 16 个是欧盟国家)的大学、研究机构和工业界的 87 个组织将通力合作,用计算机模

拟的方法研究人类大脑是如何工作的。该研究有望促进人工智能、机器人和神经形态计算系统的发展，奠定医学进步的科学和技术基础，有助于神经系统及相关疾病的诊疗及药物测试。

人脑计划旨在探索和理解人脑运行过程，研究人脑的低能耗、高效率运行模式及其学习功能、联想功能、创新功能等，通过信息处理、建模和超级计算等技术开展人脑模拟研究，为通过超级计算技术开展人脑诊断和治疗、人脑接口和人脑控制机器人研究以及开发类似人脑的高效节能超级计算机等[455]。

人脑计划分为五个方面，每个方面都是以现有工作为基础，进一步开展研究[309]。

### 1. 数据

采集筛选过的、必要的战略数据来绘制人脑图谱并设计人脑模型，同时吸引项目外的研究机构来贡献数据。当今的神经认知科学已经积累了海量实验数据，大量原创研究带来了层出不穷的新发现。即便如此，构建多层次大脑图谱和统一的大脑模型所需的绝大部分核心知识依然缺失。因此，人脑计划的首要任务是采集和描述筛选过的、有价值的战略数据，而不是进行漫无目的的搜寻。人脑计划制定了数据研究的三个重点。

(1) 老鼠大脑的多层级数据。此前研究表明，对老鼠大脑的研究成果同样适用于所有的哺乳类动物。因此，对老鼠大脑组织的不同层级间关系的系统研究将会为人脑图谱和模型提供关键参考。

(2)人脑的多层级数据。老鼠大脑的研究数据在一定程度上可以为人脑研究提供重要参考，但显然两者存在根本区别。为了定义和解释这些区别，人脑计划的研究团队采集关于人类大脑的战略数据，并尽可能积累到已有的老鼠大脑数据的规模，便于对比。

(3) 人脑认知系统结构。弄清大脑结构和大脑功能之间的联系是 HBP 的重要目标之一。HBP 会把三分之一的研究重点放在负责具体认知和行为技能的神经元结构上，从其他非人类物种同样具备的简单行为一直到人类特有的高级技能，例如语言。

### 2. 理论

人脑研究的数学和理论基础。定义数学模型，解释不同大脑组织层级与它们在实现信息获取、信息描述和信息储存功能之间的内在关系。如果缺乏统一、可靠的理论基础，我们很难解决神经科学在数据和研究方面碎片化的问题。因此，HBP 应包含一个专注于研究数学原理和模型的理论研究协调机构，这些模型用来解释大脑不同组织层级与它们在实现信息获取、信息描述和信息储存功能之间的内在关系。作为这个协调机构的一部分，人脑计划应建立一个开放的“欧洲理论神经科学研究机构”(European Institute for Theoretical Neuroscience)，以吸引更多项目外的优秀科学家参与其中，并充当创新性研究的孵化器。

### 3. 信息与通信技术平台

建立一套综合的信息与通信技术平台(Information and Communications Technology Platforms，ICT)系统，为神经认知学家、临床研究者和技术开发者提供服务以提高研究效率。建议组建六大平台，即神经信息系统、人脑模拟系统、高性能计算系统、医疗信息系统、神经形态计算系统和神经机器人学系统。

(1) 神经信息系统。人脑计划的神经信息平台将为神经科学家提供有效的技术手段，

使他们更加容易对人脑结构和功能数据进行分析,并为绘制人脑的多层级图谱指明方向。此平台还包含神经预测信息学的各种工具,这有助于对描述大脑组织不同层级间的数据进行分析并发现其中的统计性规律,也有助于对某些参数值进行估计,而这些值很难通过自然实验得出。在此前的研究中,数据和知识的缺乏往往成为我们系统认识大脑的一个重要障碍,而上述技术工具的出现使这一难题迎刃而解。

(2) 人脑模拟系统。人脑计划会建立一个足够规模的人脑模拟平台,旨在建立和模拟多层次、多维度的人脑模型,以应对各种具体问题。该平台将在整个项目中发挥核心作用,为研究者提供建模工具、工作流和模拟器,帮助他们从老鼠和人类的大脑模型中汇总出大量且多样的数据来进行动态模拟。这使"计算机模拟实验"成为可能,而在只能进行自然实验的传统实验室中是无法做到这一点的。借助平台上的各种工具可以生成各种输入值,而这些输入值对于人脑计划中的医学研究(疾病模型和药物效果模型)、神经形态计算、神经机器人研究至关重要。

(3) 高性能计算系统。人脑计划的超级计算平台将为建立和模拟人脑模型提供足够的计算能力。其不仅拥有先进的百亿亿次级超级计算技术,还具备全新的交互计算和可视化性能。

(4) 医疗信息系统。人脑计划的医疗信息系统需要汇集来自医院档案和私人数据库的临床数据(以严格保护病人信息安全为前提)。这些功能有助于研究者定义出疾病在各阶段的"生物签名",从而找到关键突破点。一旦研究者拥有了客观的、有生物学基础的疾病探测和分类方法,他们将更容易找到疾病的根本起源,并相应地研发出有效治疗方案。

(5) 神经形态计算系统。人脑计划的神经形态计算平台将为研究者和应用开发者提供他们所需的硬件和设计工具来帮助他们进行系统开发,同时还会提供基于大脑建模多种设备及软件原型。借助此平台,开发者能够开发出许多紧凑的、低功耗的设备和系统,而这些正在逐渐接近人类智能。

(6) 神经机器人平台。人脑计划的神经机器人平台为研究者提供开发工具和工作流,使他们可以将精细的人脑模型连接到虚拟环境中的模拟身体上,而以前他们只能依靠人类和动物的自然实验来获取研究结论。该系统为神经认知学家提供了一种全新的研究策略,帮助他们洞悉隐藏在行为之下的大脑的各种多层级的运作原理。从技术角度来说,该平台也将为开发者提供必备的开发工具,帮助他们开发一些有接近人类潜质的机器人,而以往的此类研究由于缺乏这个"类大脑"化的中央控制器,这个目标根本无法实现。

### 4. 应用

人脑计划的第四个主要目标是可以成功地体现出为神经认知科学基础研究、临床科研和技术开发带来的各种实用价值。

(1) 统一的知识体系原则。本项目中的"人脑模拟系统"和"神经机器人平台"会对负责具体行为的神经回路进行详尽解释,研究者可利用它们来实施具体应用,例如模拟基因缺陷的影响、分析大脑不同层级组织细胞减少的后果,建立药物效果评价模型。最终得到一个可以将人类与动物从本质上区分开来的人脑模型,例如,该模型可以表现出人类的语言能力。这些模型将使我们对大脑的认识发生质的变化,并且可以立即应用于具体的医疗和技术开发领域。

(2) 对大脑疾病的认识、诊断和治疗。研究者可充分使用医疗信息系统、神经形态

计算系统和人脑模拟系统来发现各种疾病演变过程中的生物签名，并对这些过程进行深入分析和模拟，最终得出新的疾病预防和治疗方案。这项工作将充分体现出 HBP 项目的实用价值。新诊断技术在疾病还未造成不可逆的危害前，就能提前对其进行诊断，并针对每位患者的实际情况研发相应的药物和治疗方案，实现“个人定制医疗”，这将最终有利于患者治疗并降低医疗成本。对疾病更好的了解和诊断也会优化药物研发进程，更好地筛选药物测试候选人和临床测试候选人，这无疑有益于提高后期的实验成功率，降低新药研发成本。

(3) 未来计算技术。研究者可以利用人脑计划的高性能计算系统、神经形态计算系统和神经机器人平台来开发新兴的计算技术和应用。高性能计算平台将会为他们配备超级计算资源，以及集成了多种神经形态学工具的混合技术。借助神经形态计算系统和神经机器人平台，研究者打造出极具市场应用潜力的软件原型。这些原型包括家庭机器人、制造机器人和服务机器人，它们虽然看起来不显眼，但却具备强大的技术能力，包括数据挖掘、机动控制、视频处理和成像以及信息通信等。

## 5. 社会伦理

考虑到人脑计划的研究和技术带来的巨大影响，该项目会组建一个重要的社会伦理小组，以资助针对人脑计划项目对社会和经济造成的潜在影响的学术研究，该小组会在伦理观念上影响人脑计划研究人员，管理和提升他们的伦理道德水平和社会责任感，其首要任务是在具有不同方法论和价值观的利益相关者和社会团体之间展开积极对话。

人脑计划的路线图见图 16.2[797]。

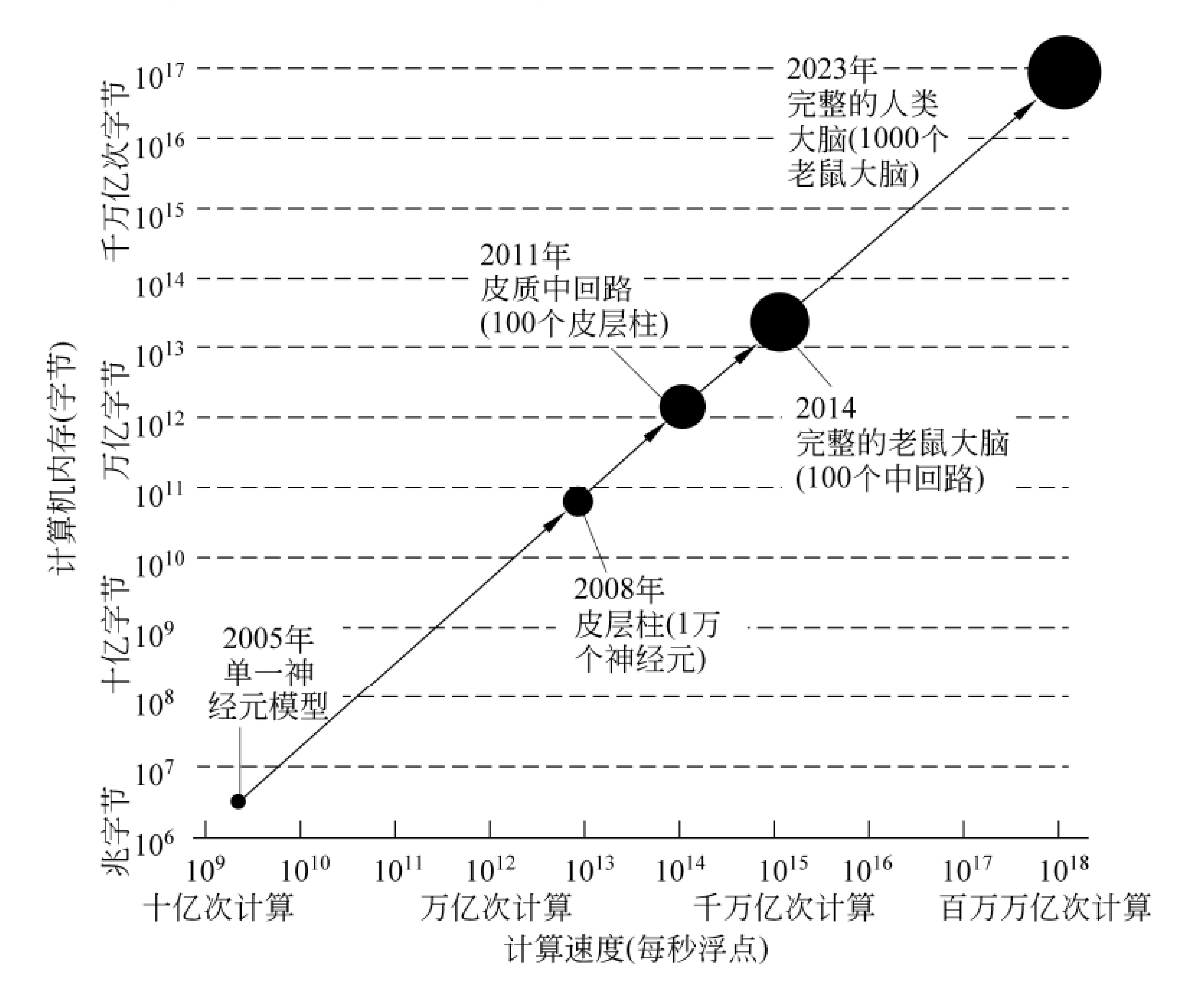



图 16.2 人脑计划的路线图

## 16.5 美国脑计划

2013年4月2日,美国白宫正式宣布"通过推动创新性神经技术进行脑研究(Brain Research through Advancing Innovative Neurotechnologies, BRAIN)"的计划,简称"脑计划"。该计划被认为可与人类基因组计划相媲美,以探索人类大脑工作机制,绘制脑活动全图,针对无法治愈的大脑疾病开发新疗法。

美国"脑计划"公布后,国家卫生研究院随即成立"脑计划"工作组。"脑计划"工作组提出了9个资助领域:统计大脑细胞类型;建立大脑结构图;开发大规模神经网络记录技术;开发操作神经回路的工具;了解神经细胞与个体行为之间的联系;把神经科学实验与理论、模型、统计学等整合;描述人类大脑成像技术的机制;为科学研究建立收集人类数据的机制;知识传播与培训。

人脑图谱是21世纪科学的极大挑战。人脑连接体项目(Human Connectome Project, HCP)将阐明大脑功能和行为背后的神经通路,是应对这一挑战的关键因素。

该研究项目(WU-Minn HCP Consortium)由华盛顿大学、明尼苏达大学、牛津大学领导,其目标是使用无创性影像学的尖端技术,创建1200个健康成人(双胞胎和他们的非孪生兄弟姐妹)的综合人脑回路图谱,将会产生大脑连通性的宝贵信息,揭示与行为的关系,遗传和环境因素对大脑行为个体差异的贡献[777]。华盛顿大学的范·埃森(D van Essen)实验室开发连接组工作台,这将提供灵活、方便用户访问、免费存储在ConnectomeDB数据库的海量数据,并在开发其他脑图谱分析方法方面发挥带头作用。连接组工作台的beta版本已经发布在网站www.humanconnectome.org上。

美国波士顿大学认知和神经系统学院长期开展脑神经模型的研究。早在1976年,格罗斯伯格(S Grossberg)就提出了自适应共振理论(ART)[287,288]。自顶向下期望控制预测性编码和匹配,以此有利于集中注意力,使同步化和增益调节注意特征表象,并且引发能有效抵制彻底遗忘的快速学习。实现快速稳定学习而不致彻底遗忘的目标通常被归结为稳定性/可塑性两难问题[290]。稳定性/可塑性两难问题是每一个需要快速而且稳定地学习的脑系统必须要解决的问题。如果脑系统设计太节省,那么我们应当希望找到一个在所有脑系统中运行的相似的原理,这个原理可以基于整个生命过程中不断变化的条件做出不同的反应来稳定学习不断增长的知识。ART预设人类和动物的感知和认知的一些基本特征就是解决大脑稳定性/可塑性两难问题的部分答案。尤其是,人类是一种有意识的生物,可以学习关于世界的预期并且对将要发生的事情做出推断。人类还是一种注意力型的生物,会将数据处理的资源集中于任何时候有限数量的可接收信息上。人类怎么会既是有意识的又是注意型的生物?这两种处理程序是相关联的吗?稳定性/可塑性两难问题以及运用共振状态的解决方案提供了一种理解这个问题的统一框架。

ART假设在使得我们快速而稳定地学习这个不断变化世界这一过程的机制,以及使得我们学习关于这个世界的推测、验证关于它的假设和将注意力集中于我们感兴趣的信息上这一过程的机制之间有密切的联系。ART还提出,要解决稳定性/可塑性两难问题,只有共振状态可以驱动快速的新学习过程,这也是这个理论名称的由来。

最近的ART模型称为LAMINART,开始展示ART的预测可能在丘脑皮质回路中得

以具体化[293]。LAMINART 模型使得视觉发展、学习、感知组织、注意和三维视觉的性质一体化。然而,它们没有将学习的峰电位动力学、高阶特异性丘脑核和非特异性丘脑核、规律性共振和重置的控制机制,以及药理学调制包含在内。

2008 年,格罗斯伯格等提出了同步匹配适应共振理论 SMART(Synchronous Matching Adaptive Resonance Theory)模型[295],分析大脑如何协调多级的丘脑和皮质进程来快速学习、稳定记忆外界的重要信息。同步匹配适应共振理论 SMART 模型,展示了自底向上和自顶向下的通路是如何一起工作并通过协调学习、期望、专注、共振和同步这几个进程来完成上述目标的。特别地,SMART 模型解释了怎样通过大脑细微回路,尤其是在新皮层回路中的细胞分层组织实现专注学习的需求,以及它们是怎样和第一层(如外侧膝状体(Lateral Geniculate Nucleus,LGN))、更高层(如枕核)以及非特异性丘脑核相互作用的。

SMART 模型超越 ART 和 LAMINART 模型的地方在于说明了这些特征怎样自然地在 LAMINART 结构中共存。特别是 SMART 解释和模拟了浅层皮质回路可能是怎样与特异性初级和较高级丘脑核以及非特异性丘脑核相互作用,从而控制用于调控认知学习和抵制彻底遗忘的动态缓冲学习记忆的匹配或不匹配的过程;峰电位动力学怎样被包含在振动频率可以提供附加的可用来控制认知导向的诸如匹配和快速学习的动作的同步共振中的;基于乙酰胆碱的过程怎样有可能使得被预测的警觉控制的性质具体化,这个性质只利用网络上本地的计算信号控制经由对不断变化的环境数据敏感的方式来学习识别类的共性规律。

SMART 模型首次从原理上将认知与大脑振动联系起来,特别是在 $\gamma$ 和 $\beta$ 频域,这是从一系列皮质和皮质下结构中得到的记录。SMART 模型表明 $\beta$ 振动为什么可以成为调制的自顶向下的反馈和重置的标志。SMART 模型发展了早前的模拟工作,解释了当调制的自顶向下的期望与连贯的自底向上的输入类型相匹配时,$\gamma$ 振动是怎样产生的。这样一个匹配使得细胞更有效地越过他们的激励阈值来激发动作电位,进而导致在共享自顶向下的激发调制的细胞中局域 $\gamma$ 频率同步的整体性增强。

SMART 模型还将不同的振动频率与峰电位时序相关的突触可塑性(Spike Timing-Dependent Plasticity,STDP)联系在一起。在突触前和突触后细胞的平均激励周期为 10～20ms 时,也就是在 STDP 学习的窗口中时,学习情景更易被限制在匹配条件下,这与实验结果相符。这个模型预测 STDP 将进一步加强相关的皮质和皮质下区域的同步兴奋度,在快速学习规律中对长期记忆权值的影响可以被匹配状态下的同步共振阻止或者快速反转。在匹配状态下被放大的 $\gamma$ 振动,通过将突触前激动压缩进狭窄的时域窗口,将有助于激动传遍皮质等级结构。这个预测与观察到的外侧膝状体成对的突触前激励对在视觉皮层中产生突触后兴奋的效果在激动间隔增加的时候快速降低是相一致的。

不同的振荡频率与匹配/共振($\gamma$ 频率)和不匹配/重置($\beta$ 频率)一起,将这些频率联系起来,不仅为选择学习,更为发现支持新学习的皮质机制的活跃的搜索过程。不匹配也预测会在 N200 ERP 的组成部分中表达的事实,指出新实验可以将 ERP 和振荡频率结合起来,作为动态规律性学习的认知过程索引。

在美国国家科学基金会的资助下,波士顿大学认知和神经系统学院成立了教育、科学和技术学习卓越中心(CELEST)。在 CELEST,计算模型的设计者、神经科学家、心理学家和工程师,与来自哈佛大学、麻省理工学院、布兰代斯大学和波士顿大学的认知和神经系统部

门的研究人员进行交流协作,研究有关脑如何计划、组织、通信、记忆等基本原理,特别是应用学习和记忆的脑模型,构建低功耗、高密度的神经芯片,实现愈来愈复杂的大规模脑回路,解决具有挑战性的模式识别问题。

波士顿大学认知和神经系统学院设计了一种软件称为模块化神经探索旅游代理(Modular Neural Exploring Traveling Agent, MoNETA)[785],它是一个芯片上的大脑。MoNETA 将运行在美国加利福尼亚惠普实验室研发的类脑(Brain Inspired)微处理器上,其工作原理正是那些把我们哺乳动物与无智商的高速机器区别开来的最基本原则。MoNETA 正好是罗马神话中记忆女神的名字"莫内塔",会做其他计算机从未做过的事情。它将感知周围的环境,决定哪些信息是有用的,然后将这些信息加入到逐渐成形的现实结构中;而且在一些应用中,它会制定计划以保证自身的生存。换句话说,MoNETA 将具有如同蟑螂、猫以及人所具有的动机。MoNETA 与其他人工智能的区别在于,它不需要显式地编程,像哺乳动物的脑一样具有适应性和效用性,可以在各种各样的环境下进行动态学习。

## 16.6 脑模拟系统 SPAUN

人类的大脑是一个高度复杂的器官,科学家们要构建出一个人工大脑模型,首先得知道我们大脑的工作原理,具体来说就是要先了解大脑里每一个部分负责的运算任务,以及这些运算功能在神经网络系统上的实现原理。2012 年 11 月,*Science* 上发表了伊莱亚史密斯(C Eliasmith)等人的文章[210],介绍一种大规模的人体大脑模型,如图 16.3 所示。这种大脑模型就能够模拟各种复杂的人类行为,这一成果标志着科学家们在人工智能研究领域又前进了一大步。

伊莱亚史密斯等人开发的大脑仿真模型称为语义指针结构统一网络(Semantic Pointer Architecture Unified Network, SPAUN),如图 16.3 所示。该系统能够观察图像,并使用配套的模型手臂做出相应的动作。伊莱亚史密斯等人开发的 SPAUN 系统可以完成八种各不相同的任务,在所有这八种任务中都会包含对各种图形(主要是数字图形)的介绍,以及根据图形做出相应的动作(用人工臂画出"看到"的数字)。这些任务包括简单的图像识别任务、记忆性的任务(按照看到数字的先后顺序重新写一遍)、强化学习任务(如赌博任务)和更加复杂的认知任务(类似于智商测试题一类的任务)。SPAUN 系统会依靠它所拥有的那 250 万个神经元细胞来完成这些测试,这些神经元细胞按照我们人类大脑的组成方式形成了多个子系统,这些子系统分别对应了我们人类大脑的不同区域,最后这些子系统之间又互相联系起来,具备了大脑最基本的功能[451]。

人工大脑看到的视觉图像信息首先会被"压缩"处理,去除掉不相关的、或者冗余的信息。伊莱亚史密斯小组在对图形信息进行压缩处理时使用的是一种多层次受限玻尔兹曼机的算法,这种算法属于一种前馈神经网络系统的运行机制,每一层有限玻尔兹曼机处理都可以得到一种图形特征信息,经过多轮(层)有限玻尔兹曼机处理之后就可以得到整个图形的所有相关信息。然后,将这些图形信息一一分配给人工大脑里与真正人类大脑视觉中枢对应的各个子系统,它们分别对应初级视觉皮层和次级视觉皮层、纹状体外皮层和颞下皮层。在运动功能方面,SPAUN 系统也采取了一种类似的方法,他们将简单的动作命令,比如画出数字"6",也分解为很多个简单的动作,然后将这些动作组合起来就可以画出一个"复杂

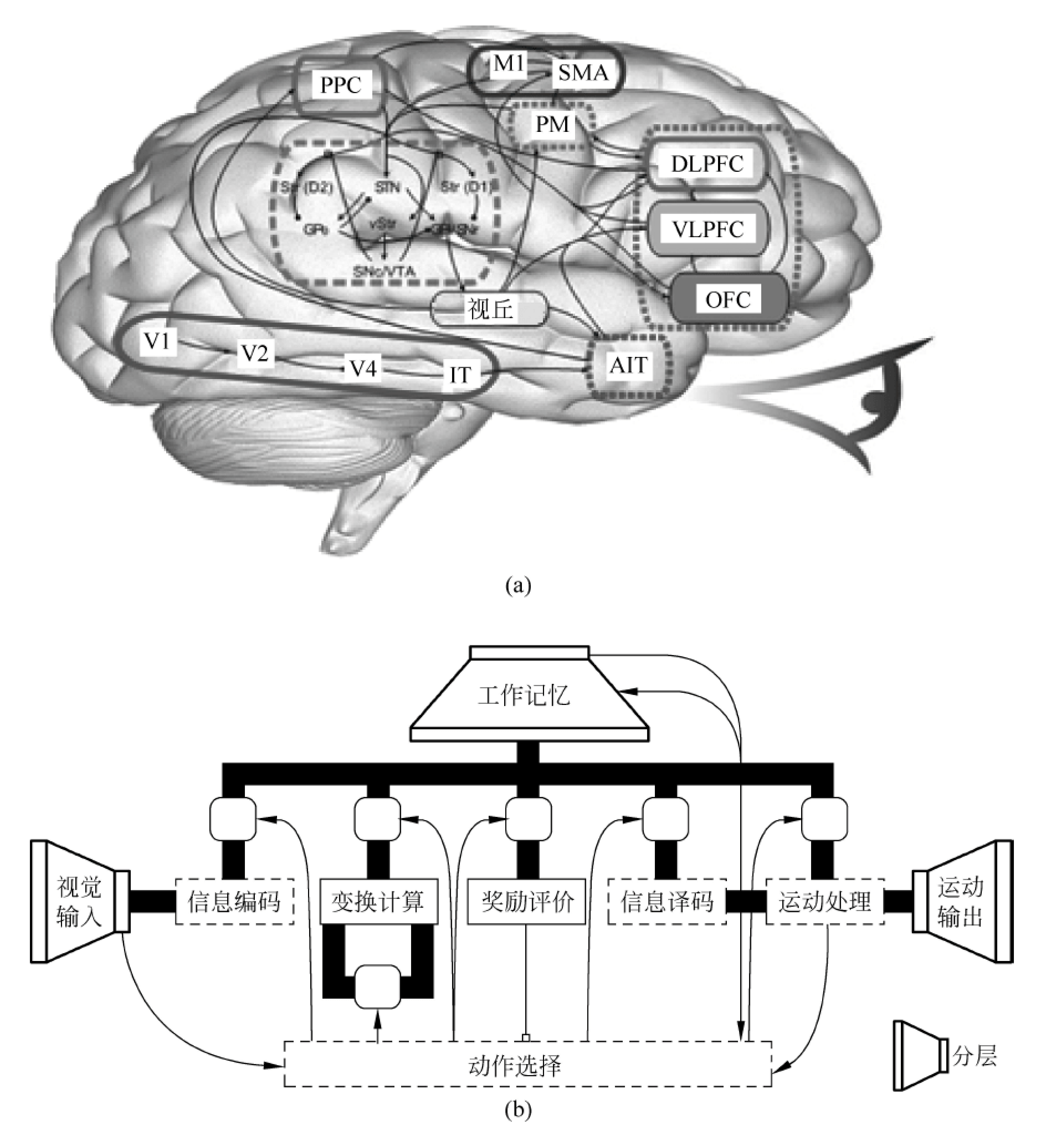



图 16.3 SPAUN 的大脑仿真模型

的"6。所有相关的运算全都基于最佳控制理论,其中还包括辅助运动中枢和初级运动中枢的运算。这种对信号的压缩处理同时伴以动作的人工大脑模型解决了大脑在与环境发生相互作用时需要处理的"广度难题",以往的人工模型在处理这类问题时总是不知道该如何处置大量的感觉信息,同时也不知道在面对众多动作备选方案时应该做出哪种选择。

SPAUN 系统的认知装置实际上包括两个相互交叉的组成部分,它们分别是相当于人类大脑前额皮质区的工作记忆系统和相当于人类大脑基底神经节和丘脑的动作选择系统[220]。这套动作选择系统控制着人工大脑当前的状态,同时也部分受到了强化学习理论和当前流行的基底神经节模型的启发。SPAUN 系统的记忆工作系统采用了一套全新的算法,这套算法借鉴了计算神经科学领域的神经系统算法和来自数学心理学领域的卷积记忆理论。这套神经系统算法使 SPAUN 系统拥有了一个网络化的信息存储机制,而卷积记忆理论又让 SPAUN 系统可以将以往的信息和最新接收的信息有效地结合在一起。所以 SPAUN 系统可以有效地做出重复行为,比如写出一列数字中的第一个数字和最后一个数字,这是其他人工大脑模型无法比拟的。

伊莱亚史密斯等人还使用了另外一个记忆工作系统来自动推测以往和当前信号之间的关系。这种自动推理功能意味着最初级的句法功能，SPAUN 系统呈现出的数字识别并再现功能预示着这种句法功能在将来的某一天一定会实现。这种再现功能与符号运算功能有着直接的联系，这种符号运算在计算机科学和联结主义理论著作里非常常见。SPAUN 系统使用这些计算方法居然还通过了最基础的智商测试考核。

在 SPAUN 系统里对应前额皮质区的那些子系统起到了连接抽象运算、符号运算和单个神经元细胞活动的作用。关于卷积记忆功能，伊莱亚史密斯等人做了一个非常有意思的预测，他们估计神经元细胞激活的速率(在单位时间内出现的动作电位的平均数量)会随着不断地连续完成记忆工作而逐渐加快。

在 SPAUN 系统的每一个模块当中，实际的信息都是通过大量被激活的神经细胞来完成处理的。而在生理条件下借助神经网络开展的高水平的运算与依靠单个神经元细胞开展的低水平的运算之间的这种联系在 SPAUN 系统中是依靠所谓的“神经工程架构(Neural Engineering Framework)”来重建的[209]，这套神经工程架构系统尤其善于在活化的神经网络里完成任意数学矢量运算(Arbitrary Mathematical Vector Operation)。这套系统假定信息会按照神经激活速度的线性被读取，然后再以非线性的方式将信息转换成神经活动功能。这样每一个子系统里处理的信息都会被分配给每一个神经元细胞，这种模式也非常符合大脑电生理研究工作得到的结论，比如我们的大脑对不同的感觉刺激(输入)信号或运动输出信号的响应速度是不一样的等。

基于伊莱亚史密斯等人的开发思路，当 SPAUN 系统在某些方面不能很好地模拟真实大脑情况时，我们一点也不感到奇怪。例如，这套系统多个部分的响应活性在好几个方面(其中包括最基础的统计范畴)都和大脑的实际情况有明显的不同。我们现在还不知道将来这些问题能够被改善到何种程度，也不清楚这些偏差在多大程度上是我们大脑内部基础响应水平不一致情况的真实反映。SPAUN 系统的最大问题还是它硬连接的本质以及不能学习新任务(功能)的特点。不过 SPAUN 系统的结构具有非常大的灵活性，并不拘泥于某一项任务，而且在 SPAUN 系统中有多个部分都是具备学习功能的，例如，图形信息多层处理系统和动作选择系统都有这种学习的潜力。至于说更广义的学习能力，如学习一项全新的任务，这也许是伊莱亚史密斯等人故意留下的一个空白。实际上，SPAUN 系统所欠缺的恰恰就是我们在对自身大脑认识上还有所不足的部分。伊莱亚史密斯等已经将大量的大脑研究成果纳入 SPAUN 系统，这件工作本身就已经向我们展现了一个大脑工作理论，当然其中并不包括与学习相关的机制。此外，伊莱亚史密斯等人也提供了一种大规模的、自上而下开发人工智能系统的可能性。SPAUN 系统的出现为这方面的工作设立了一个新的标杆，也提供了一条新的途径，不要只想着如何将尽可能多的神经细胞或信息量集中在一起，注意力应该集中在尽可能地重现大脑功能、行使更复杂的行为上。

## 16.7 神经形态芯片

计算机的“冯·诺依曼架构”与“人脑架构”的本质结构不同，人脑的信息存储和处理，通过突触这一基本单元来实现，因而没有明显的界限。正是人脑中的千万亿个突触的可塑性——各种因素和各种条件经过一定的时间作用后引起的神经变化(可变性、可修饰性等)，

使得人脑的记忆和学习功能得以实现。

模仿人类大脑的理解、行动和认知能力,成为重要的仿生研究目标,该领域的最新成果就是推出了神经形态芯片。《麻省理工科技评论》(*MIT Technology Review*)2014 年 4 月 23 日刊出了“2014 十大突破性科学技术”的文章,高通(Qualcomm)公司的神经形态芯片(Neuromorphic Chips)名列其中。

### 16.7.1 神经形态芯片简史

神经形态芯片的研究已有 20 多年的历史。1989 年,加州理工学院米德(C Mead)在文献[486]中给出了神经形态芯片的定义:“模拟芯片不同于只有二进制结果(开/关)的数字芯片,可以像现实世界一样得出各种不同的结果,可以模拟人脑神经元和突触的电子活动。”然而,米德本人并没有完成模拟芯片的设计。

语音处理芯片公司 Audience 公司,对神经系统的学习性和可塑性、容错、免编程、低能耗等特征进行了研究,研发出基于人的耳蜗而设计的神经形态芯片,可以模拟人耳抑制噪声,应用于智能手机。Audience 公司也由此成为行业内领先的语音处理芯片公司。

高通公司的“神经网络处理器”与一般的处理器工作原理不同。从本质上讲,它仍然是一个由硅晶体材料构成的典型计算机芯片,但是它能够完成“定性”功能,而非“定量”功能。高通开发的软件工具可以模仿大脑活动,处理器上的“神经网络”按照人类神经网络传输信息的方式而设计,它可以允许开发者编写基于“生物激励”程序。高通设想其“神经网络处理器”可以完成“归类”和“预测”等认知任务。

高通公司给其“神经网络处理器”起名为“Zeroth”。Zeroth 的名字起源于“第零原则”。“第零原则”规定,机器人不得伤害人类个体,或者因不作为致使人类个体受到伤害。高通公司研发团队一直致力于开发一种突破传统模式的全新计算架构。他们希望打造一个全新的计算处理器,模仿人类的大脑和神经系统,使终端拥有大脑模拟计算驱动的嵌入式认知——这就是 Zeroth。“仿生式学习”“使终端能够像人类一样观察和感知世界”“神经处理单元(NPU)的创造和定义”是 Zeroth 的三个目标。关于“仿生式学习”,高通公司是通过基于神经传导物质多巴胺的学习(又名“正强化”)完成的,而非编写代码实现。

IBM 公司自 1956 年创建第一台人脑模拟器(512 个神经元)以来,就一直在从事对类脑计算机的研究,模仿了突触的线路组成、基于庞大的类神经系统群开发神经形态芯片也就自然而然地进入了其视野。其中,IBM 第一代神经突触(Neurosynaptic)芯片用于“认知计算机”的开发。尽管“认知计算机”无法像传统计算机一样进行编程,但可以通过积累经验进行学习,发现事物之间的相互联系,模拟大脑结构和突触可塑性。

2008 年,在美国国防高级研究计划局(DARPA)的资助下,IBM 的“自适应可变神经可塑可扩展电子设备系统”项目(SyNAPSE)第二阶段项目则致力于创造既能同时处理多源信息又能根据环境不断自我更新的系统,实现神经系统的学习性和可塑性、容错、免编程、低能耗等特征。项目负责人莫得哈(D Modha)认为,神经芯片将是计算机进化史上的又一座里程碑。

2011 年,IBM 首先推出了单核含 256 个神经元,256×256 个突触和 256 个轴突的芯片

原型。当时的原型已经可以处理像玩 Pong 游戏这样复杂的任务。不过相对来说还是比较简单,从规模上来说,这样的单核脑容量仅相当于虫脑的水平。2013 年 8 月初,IBM 研究院公布了类脑系统 TrueNorth 计算机系统架构和编程语言,希望在某些应用场景下取代今天的计算机。

2014 年 8 月 8 日,IBM 在 *Science* 刊物上公布仿人脑功能的芯片[495],能够模拟人脑神经元、突触功能以及其他脑功能,从而完成计算功能,这是模拟人脑芯片领域所取得的又一大进展。IBM 表示,这款名为 TrueNorth 的微芯片擅长完成模式识别和物体分类等烦琐任务,而且功耗还远低于传统硬件。

2003 年,英国 ARM 公司开始研制类脑神经网络的硬件单元,称为 SpiNNaker(Spiking Neural Networks Architecture)。2011 年,正式发布了包含 18 个 ARM 核的 SpiNNaker 芯片。2013 年,开发基于 UDP 的 Spiking 接口,可以用于异质神经形态系统的通信。2014 年,将与滑铁卢大学合作,支持 Spaun 模型的硬件计算。

2011 年,德国海德堡大学在 FACTS 项目的基础上,在 Proactive FP7 的资助下,启动了为期 4 年的 BrainScales 项目[656]。2013 年,加入欧盟"人类大脑计划"。在 2013 年 6 月 20 日结束的莱比锡世界超级计算机大会上,人脑研究项目协调人之一,德国海德堡大学教授麦耶(Meier K)介绍了德国科学家取得的研究进展。麦耶宣布,神经形态系统将出现在硅芯片上或硅圆片上。这不仅是一种芯片,而且是一个完整的硅圆片,上面集成了 20 万个神经元和 500 万个突触。这个硅圆片的大小如同一个略大些的盘子。这些硅圆片就是未来 10 年欧盟人脑研究项目要开发的类似人脑的新型计算机系统结构的基石。

中国科学院计算技术研究所陈天石、陈云霁等于 2012 年提出了国际上首个人工神经网络硬件的基准测试集 benchNN[137]。这项工作提升了人工神经网络处理速度,有效加速了通用计算。先后推出了一系列寒武纪神经网络专用处理器:DianNao(面向多种人工神经网络的原型处理器结构)、DaDianNao(面向大规模人工神经网络)和 PuDianNaoPuDianNao(面向多种机器学习算法)等。

### 16.7.2 IBM 的 TrueNorth 神经形态系统

经过 3 年的努力,IBM 在复杂性和使用性方面取得了突破。2014 年 8 月 8 日,IBM 在 *Science* 刊物上公布仿人脑功能的 TrueNorth 的微芯片[495],如图 16.4 所示。这款芯片能够模拟神经元、突触的功能以及其他脑功能执行计算,擅长完成模式识别和物体分类等烦琐任务,而且功耗还远低于传统硬件。该芯片由三星电子负责生产,拥有 54 亿个晶体管,是传统 PC 处理器的 4 倍以上。它的核心区域内密密麻麻地挤满了 4096 个处理核心,产生的效果相当于 100 万个神经元和 2.56 亿个突触。目前,IBM 已经使用 16 块芯片开发了一台神经突触超级计算机。

TrueNorth 的 4096 个核心之间就使用了类似于人脑的结构,每个核心包含约 120 万个晶体管,其中负责数据处理和调度的部分只占用少量晶体管,而大多数晶体管都被用于数据存储以及与其他核心沟通等方面。在这 4096 个核心中,每个核心都有自己的本地内存,它们还能通过一种特殊的通信模式与其他核心快速沟通,其工作方式非常类似于人脑神经元

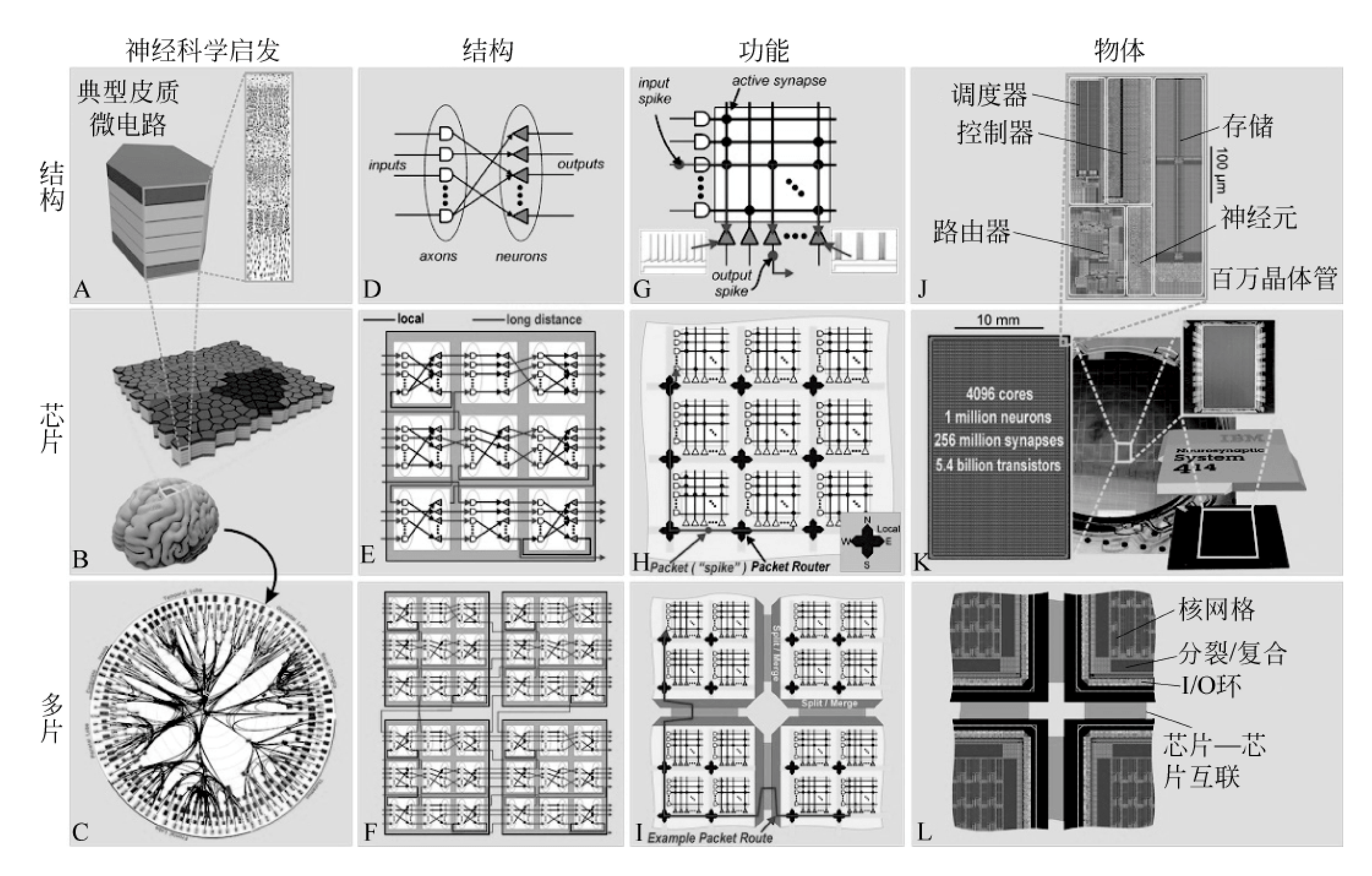



图 16.4 IBM 的 TrueNorth 芯片

与突触之间的协同，只不过，化学信号在这里变成了电流脉冲。IBM 把这种结构称为“神经突触内核架构”。

IBM 使用软件生态系统将众所周知的算法，例如，卷积网络、液态机器、受限玻尔兹曼机、隐马尔科夫模型、支持向量机、光学流量、多模态分类通过离线学习加到系统结构中。现在这些算法在 TrueNorth 中运行无须改变。为了测试在现实世界中的应用问题，开发了固定相机配置的多目标检测和分类应用。该任务有两个挑战。

(1) 在稀疏图像下准确检测人群、自行车、汽车、卡车和公交车。

(2) 正确辨别物体。

在 400 像素×240 像素孔径下操作，每秒 30 帧 3 色视频(见图 16.5)，芯片耗电 63mW。跟传统计算机用 FLOPS(每秒浮点运算次数)衡量计算能力一样，IBM 使用 SOP(每秒突触运算数)来衡量这种计算机的能力和能效。其完成 460 亿 SOP 所需的能耗仅为 1W。

通信效率极高，从而大大降低能耗这是这款芯片最大的卖点。TrueNorth 的每一内核均有 256 个神经元，每一个神经有分别都跟内外部的 256 个神经元连接。

但是相比之下，人脑有上千亿个神经元，每个神经元又有成千上万的突触，那样一个神经网络就更加无法想像了。IBM 的最终目标就是希望建立一台包含 100 亿个神经元和 100 万亿个突触的计算机，这样的计算机要比人类大脑的功都强大 10 倍，而功耗只有 1000W。

## 16.7.3 英国 SpiNNaker

SpiNNaker 是曼彻斯特、南安普顿、剑桥、谢菲尔德等地多所大学和企业机构联合发起的项目，并得到了英国工程和自然科学研究委员会(EPSRC)500 万英镑的投资。负责领衔

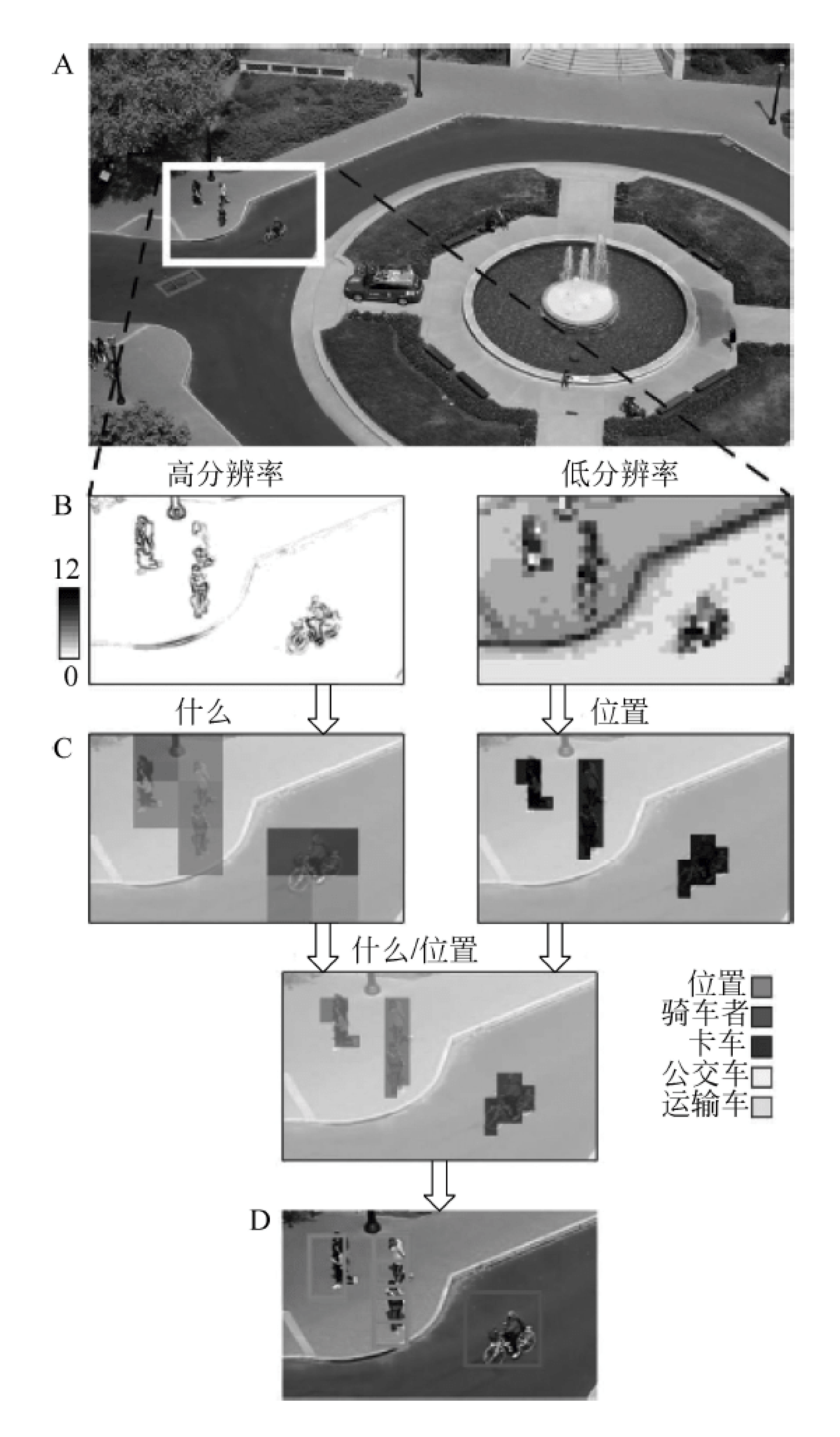



图 16.5　多目标检测和分类

的是曼彻斯特大学教授弗伯(S Furber),他从事人脑功能与架构研究很多年,同时也是 ARM 处理器核心鼻祖 Acorn RISC Machine 的联合设计师之一。这一项目在 2005 年获得批准之后,ARM 公司立即投入了大力支持,向科研团队提供了处理器和物理 IP。

人脑中有大约 1000 亿个神经元和多达 1000 万亿个连接,即使是一百万颗处理器也只能模拟人脑的 1%。神经元彼此通过模拟电子峰电位脉冲的方式传递信息,SpiNNaker 则利用描述数据包的方式模拟,并建立虚拟神经元。使用封包的电子数据意味着 SpiNNaker 能够以更少的物理连接像人脑那样快速传递峰电位脉冲。2011 年正式发布了包含 18 个 ARM 核的芯片[547],最新实现了 48 个节点的 PCB 板。

单个 SpiNNaker 多处理器芯片含有 18 个低功耗的 ARM 968 核,每个核可以模拟 1000 个神经元。每个芯片还有 128MB 的低功耗 SDRAM,存储神经元间的突触连接权值、突触延时等信息。单个芯片的功耗不超过 1W。芯片内采用局部同步,芯片间采用全局异步的方式。

SpiNNaker 系统没有中央计时器,这就意味着,信号的发出和接收不会经过任何时间同步,这些信号将会相互干扰,输出结果也会随着数百万微小的随机变化因素而发生改变。这

听起来似乎会造成混乱，对于数学计算等对精度要求很高的任务来说也确实如此，但是对于那些模糊运算任务，如何时松开手以便丢出一个球，或者用哪个词作为一个句子的结尾，这一系统就能从容应付，毕竟大脑在处理这类任务时不会被要求将计算结果精确到小数点后10 位，人脑更像是一个混沌系统。大量的 SpiNNaker 处理器通过以太网连接异步互联。

每个 SpiNNaker 中含有一个特制的路由器，用于完成 SpiNNaker 内部神经元间及芯片间神经元通信。Core 间采用地址时间表示通信协议进行神经动作电位时间信息传输。

2013 年，开发基于 UDP 用户数据报协议的锋电位接口，可以用于异质神经形态系统的通信。演示了 SpiNNaker 和海德堡大学 BrainScaleS 系统的混合通信，发展大规模的神经形态网络。

### 16.7.4　寒武纪神经网络处理器

随着社会从信息时代过渡到智能时代，人工智能芯片将是支撑智能计算不可或缺的载体。复杂的深度学习网络计算需求很高，需要有更多更强大的计算资源。GPU 是作为目前主流的人工智能计算平台，由于其基本框架结构并不是为人工智能所设计的，因此效率受到很多限制。而 FPGA（现场可编程门阵列）虽然迭代快，但从计算速度和能耗比来说，和专用的人工智能芯片相比仍然有差距。

由于定制化、低功耗等好处，ASIC 正在被越来越广泛地采用。谷歌为 TensorFlow 设计的的 TPU 芯片就是采用 ASIC。智能发展到现在，算法上的进步很多，也能解决很多实际应用中的问题，如模式识别。但这和人们所期望的智能还存在很大的距离。神经网络芯片对于人工智能进步，尤其是对于高级智能能力的实现，会起关键的作用。每个时代都有其核心的物质载体，如工业时代的蒸汽机、信息时代的通用 CPU，智能时代也将会出现这个核心载体。中国科学院计算技术研究所陈云霁、陈天石等研究寒武纪神经网络芯片，能在计算机中模拟神经元和突触的计算，对信息进行智能处理，还通过设计专门的存储结构和指令集，每秒可以处理 160 亿个神经元和超过 2 万亿个突触，功耗却只有原来的十分之一。未来甚至有希望把类似 AlphaGo 的系统都装进手机。

当前寒武纪系列已包含三种处理器结构：DianNao、DaDianNao 和 PuDianNao。在 2015 ACM/IEEE 计算机体系结构国际会议上，发布了第四种结构：面向卷积神经网络的 ShiDianNao。

陈天石等提出的 DianNao 是寒武纪系列的第一个原型处理器结构[137]，包含一个处理器核，主频为 0.98GHz，峰值性能达每秒 4520 亿次神经网络基本运算（如加法、乘法等），65nm 工艺下功耗为 0.485W，面积 3.02 $mm^2$（如图 16.6 所示）。在若干代表性神经网络上的实验结果表明，DianNao 的平均性能超过主流 CPU 核的 100 倍，面积和功耗仅为 CPU 核的 1/30～1/5，效能提升达三个数量级；DianNao 的平均性能与主流通用图形处理器（NVIDIAK20M）相当，但面积和功耗仅为后者的百分之一量级。

DianNao 要解决的核心问题是如何使有限的内存带宽满足运算功能部件的需求，使运算和访存之间达到平衡，从而实现高效能比。为此提出了一套基于机器学习的处理器性能建模方法，并基于该模型最终为 DianNao 选定了各项设计参数，在运算和访存间实现了平

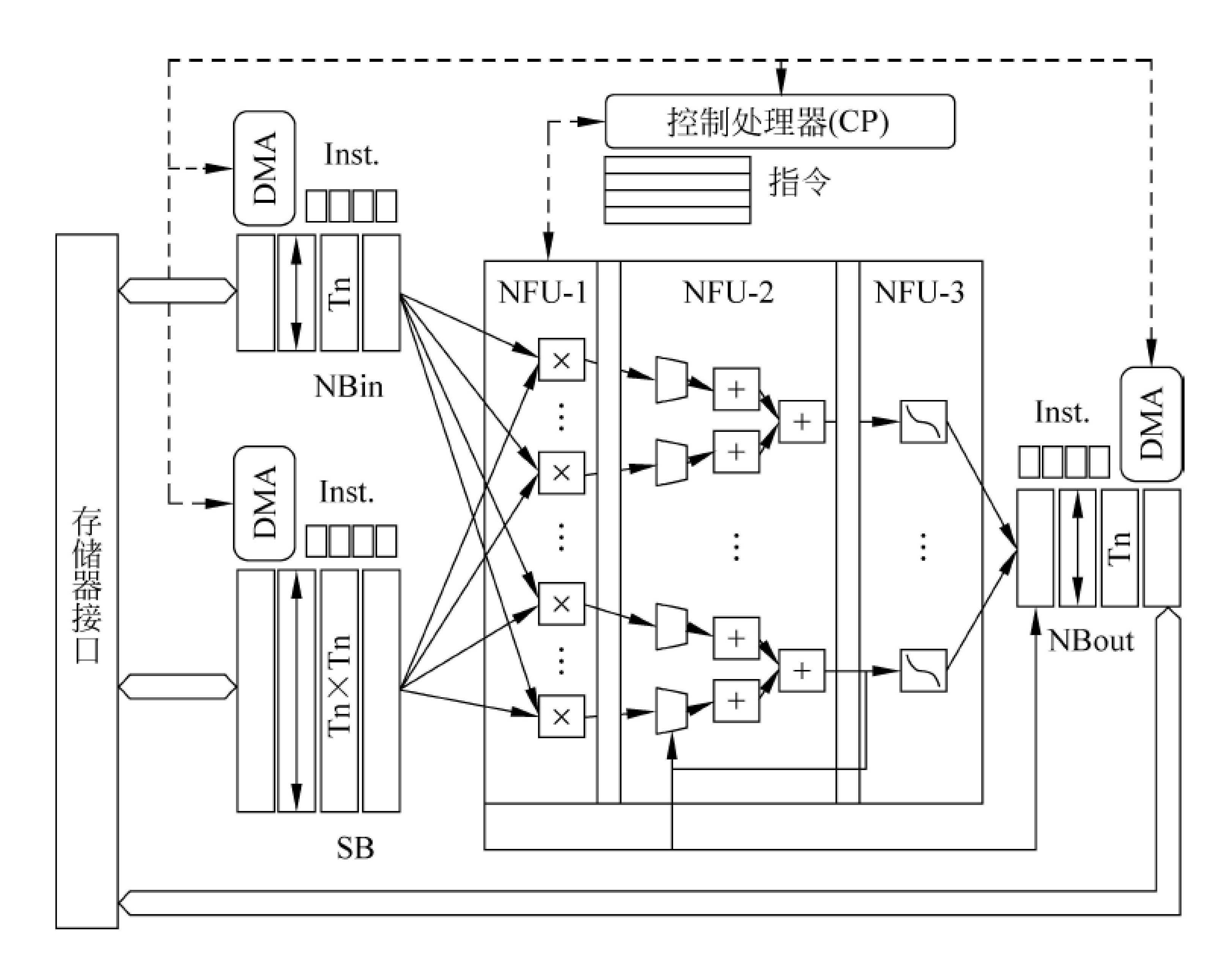



图 16.6 DianNao 结构图

衡,显著提升了执行神经网络算法时的效能。

文献[448]介绍 Cambricon 处理器,提出新的加速神经网络处理的指令集,把 Add/Multplication 这些操作都集中在了 Vector/Matrix Func Unit 里,片上存储体系的设计也更为通用,通过更为精巧的 crossbar scratchpad memory 来统一提供片上访存支持。在指令集的层次为更丰富的应用类型提供了支持。文献[837]介绍了稀疏深度学习处理器 Cambricon-X,对于稀疏连接,将访存逻辑定制在硬件层面,通过引入一个称为 IM(Index Module)的硬件模块,完成稀疏访存的处理,从而将稀疏向量/矩阵运算转换成常规向量/矩阵运算。在 Cambricon-X 里,支持的还是权重的稀疏化,对于神经元的稀疏化并没有进行特殊的支持。全连接层的神经元稀疏化可以比较自然的得到支持,而卷积层的神经元稀疏化则不太容易支持。Cambricon-X 的架构与 Cambricon 很相似,主要区别在于 Cambricon-X 中间多出了一个 Buffer Controller 模块。这个模块就是完成稀疏访存操作的核心模块。

## 16.8 脑机融合

生物智能(脑)与机器智能(机)的融合乃至一体化,将脑的感知和认知能力与机器的计算能力完美结合,有望产生令现有生物智能系统和机器智能系统均望尘莫及的更强智能形态,这种形态称为脑机融合(Brain-Computer Integration)。

### 16.8.1 脑机接口

脑机接口(Brain-Computer Interface,BCI)不依赖于大脑的正常输出通路(即外围神经和肌肉组织),就可以实现人脑与外界(计算机或其他外部装置)直接通信的系统。1973 年,

维达(Jacques Vidal)发表了第一篇关于脑机接口技术的文章[787]。

麻省理工学院、贝尔实验室和神经信息学研究所的科学家已经成功研制了一个可以模拟人类神经系统的电脑微芯片，并成功地植入大脑，利用仿生学的原理对人体神经进行修复。它与大脑协作发出复杂的指令给电子装置，监测大脑的活动都取得了很好的效果。剑桥大学的翰福瑞斯认为，在不久的将来，人们将可以在脑中放入增加记忆的微芯片，使人类有一个备用的大脑。

美国生物计算机领域的研究人员利用取自动物脑部的组织细胞与计算机硬件进行接合，这样研制而成的机器称为生物电子人或是半机械人。如果芯片与神经末梢相吻合，就可将芯片通过神经纤维和身体上脑神经系统连接起来。这样就通过计算机提高人的大脑功能。

美国南加州大学的勃格(T Berger)和列奥(J Liaw)于 1999 年提出了动态突触神经回路模型[432]，并于 2003 年研制出大脑芯片，能够代替海马功能。大脑芯片在活体小白鼠上实验成功，证明该回路模型与活体鼠脑中的信息处理是一致的。该项目是美国心智-机器合成(Mind-Machine Merger)计划的一部分，其研究成果取得了突破性进展，曾被中国科学界评为 2003 年世界十大科技进展之一。

## 16.8.2　脑机融合的认知模型

脑机融合是一种基于脑机接口技术，综合利用生物智能和机器智能的新型智能系统。脑机融合是脑机接口技术发展的必然趋势。在脑机融合系统中，大脑与大脑、大脑与机器之间不仅是信号层面上的脑机互通，更需实现大脑的认知能力与机器的计算能力的融合。但大脑的认知单元与机器的智能单元具有不同的关联关系和逻辑通路，因此，脑机融合的关键科学问题之一是如何建立脑机协同的认知计算模型。

脑机融合是当前智能科学中一个活跃的研究领域。2009 年，迪乔范纳(J DiGiovanna)等[188]设计了基于强化学习的互适应脑机接口系统，利用奖惩机制调节大脑活动，机器采用强化学习算法自适应控制机械臂运动，实现了性能更为优化的机械臂运动控制；2010 年，福山雅治(O Fukayama)等[251]通过提取和分析老鼠的运动神经信号来控制一辆机械车。2011 年，尼科勒利斯(M A L Nicolelis)团队在 *Nature* 杂志上报道了一种新型的脑—机—脑信息通路的双向闭环系统[546]，在对猴子大脑神经信息进行解码的同时将猴子触觉信息转化为电刺激信号反馈到大脑，实现了脑与机的相互配合。

脑机融合系统具有三个显著特征：①对生物体的感知更加全面，包含表观行为理解与神经信号解码；②生物体也作为系统的感知体、计算体和执行体，且与系统其他部分的信息交互通道为双向；③多层次、多粒度的综合利用生物体和机器的能力，达到系统智能的极大增强[958]。

根据脑机融合的需要，提出了一种智能体模型 ABGP[713]，如图 16.7 所示。ABGP 智能体模型由感知(Awareness)、信念(Belief)、目标(Goal)和规划(Plan)四个模块组成。这种智能体模型既考虑了智能体内部的思维状态，又考虑了对外部场景的认知和交互，对智能体决策发挥重要作用。智能体模型 ABGP 是脑机协同仿真环境的核心部件。基于心智模型 CAM 和智能体模型 ABGP，提出了脑机协同的认知模型，如图 16.8 所示。

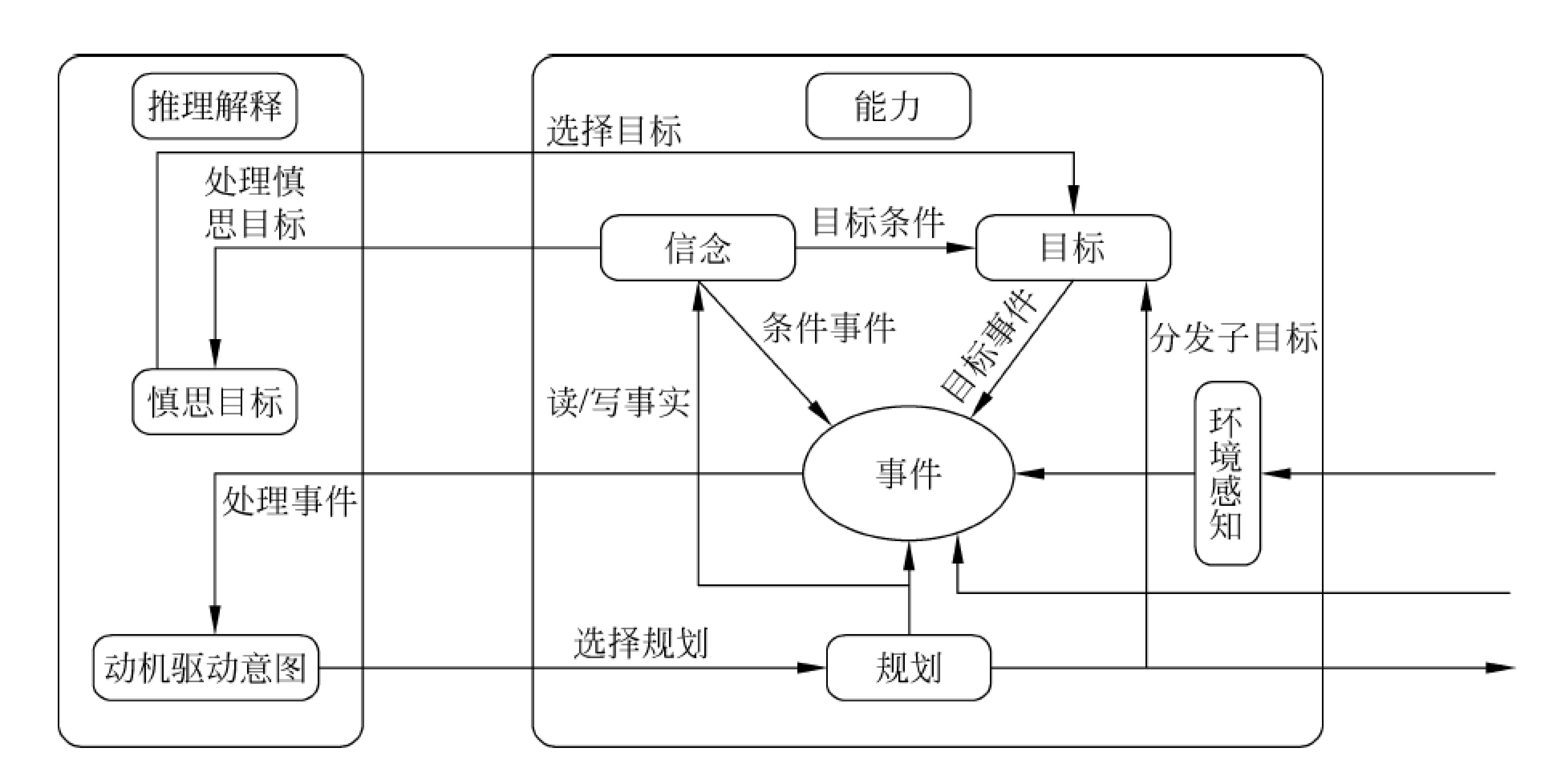



图 16.7　智能体 ABGP 模型

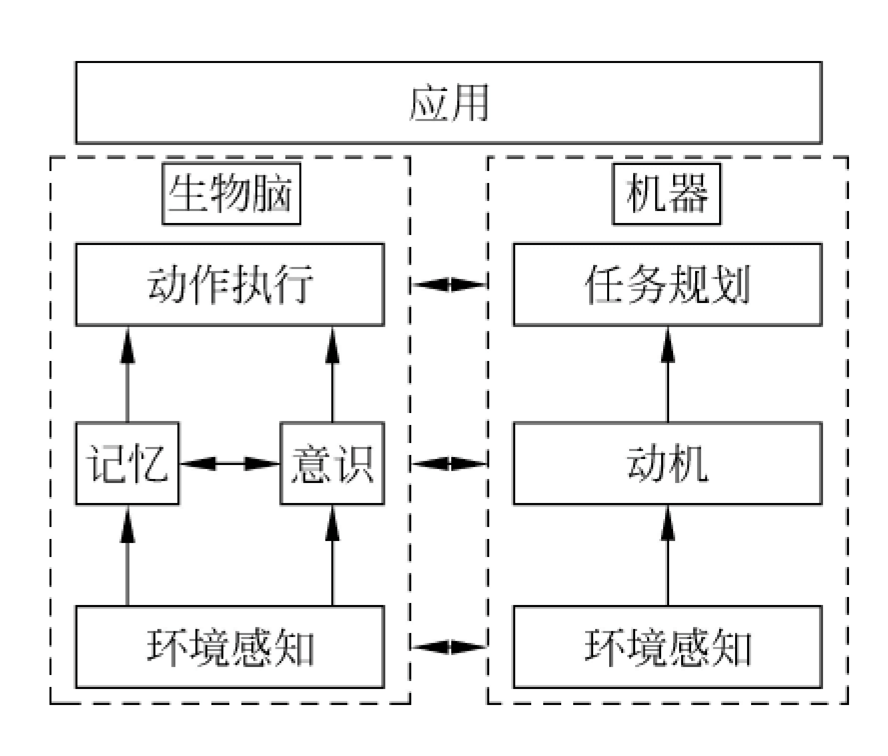



图 16.8　脑机融合的认知计算模型

### 16.8.3　脑机融合的环境感知

在环境感知方面提出了基于深度学习的视觉感知机制、基于特征整合的视觉感知方法、大数据谱聚类算法。

#### 1. 基于深度学习的视觉感知

卷积神经网络是典型的深度学习算法,在图像识别领域应用广泛。作为环境感知以CNN为基础,综合其他分类器,实现特征提取与非线性特征映射的组合。利用卷积神经网络提取不变性的视觉特征,如轮廓和边缘信息等,然后使用超限学习机完成最后的分类。

#### 2. 基于特征整合的视觉感知

感知外界环境时,人脑中形成了两类谱图:位置主谱图和特征谱图。位置主谱图记录了全局图像中每个局部底层图像特征的具体位置。特征谱图,记录了局部的底层特征的关系,称为关系编码模式。通过扫描位置主谱图,被扫描到的特征谱图被激活,形成当前物体的暂态表示。为获得对输入特征序列的整体理解,注意机制启动注意,串行扫描主谱图,通

过注意将特征联系起来。通过最大熵原则选择势函数，将特征函数联系起来，构成随机场。通过查询识别网络，随机场通过 Veterbi 算法进行连接搜索来获得对感知图像的预测，物体的底层特征和高层语义被联系起来，从而完成整个物体识别的特征捆绑。

### 3. 大数据谱聚类算法

针对谱聚类中特征分解计算复杂度过高的问题，基于自适应的 Nyström 采样方法，设计了一种自适应的 Nyström 采样方法，每个数据点的抽样概率都会在一次采样完成后及时更新。利用 Normalized Cuts 与加权核 k-means 之间的联系，设计了近似加权核 k-means 算法来优化 Normalized Cuts 的目标函数，有效降低了 Normalized Cuts 的时间和空间复杂度。

## 16.8.4 脑机融合的自动推理

动机是内部驱动力量和主观推理，直接驱使个体活动来发起并维持心理状态以达到某种特定的目的。通过动机驱动规划，实现自动推理。在脑机融合系统中，提供两种类型的动机，即基于需求的动机和基于好奇心的动机。

基于需求的动机被表示为一个 3 元组{N，G，I}，其中 N 表示需要，G 是目标，I 表示动机强度。在脑机融合系统中，需求有三类，即感知需求（Perception Needs）、适应需求（Adaptation Needs）以及合作需求（Cooperation Needs）。动机由激励规则激活。

基于好奇心的动机是通过动机学习算法建立一个新的动机。智能体将观察到的感知输入创建为一种内部表达，并且将这种表达与学习到的有利于操作的行为相联系。如果智能体的动作结果与其当前目标不相关，则不会进行动机学习，这种对学习内容的筛选是非常有用的。但是即便学习不被其他动机触发时，基于新颖性的学习仍然可以在这样的情况下发生。

动机的学习过程就是通过观察获取感知状态，然后感知状态由事件进行相互转换。发现新颖性事件激发智能体兴趣。一旦兴趣被激发，智能体的注意力可以被选择并集中在环境的一个方面。基于新颖性的动机学习算法中，利用观察函数将注意力集中在感知状态的子集，然后利用差异函数计算子集上的差异度，再借助于事件函数形成事件，事件驱动着内省搜索，利用新颖性和兴趣度选择出最感兴趣的事件项，以便让智能体专注于该事件项，最后基于所关注事件项的最大兴趣度创建一个新的动机。

## 16.8.5 脑机融合的协同决策

根据脑机协同的认知计算模型，作者提出脑机融合的协同决策[712]。脑机融合的协同决策是基于联合意图的理论，该理论可以有效地支持智能体间联合社会性行为的描述和分析。在脑机融合中，脑和机器被定义为具有共同的目标和共同的心智状态的智能体。在短时记忆支持下，采用分布式动态描述逻辑（Distributed Dynamic Description Logic，D3L）刻画联合意图。分布式动态描述逻辑充分考虑了动态描述逻辑在分布式环境下的特性，利用桥规则构成链，通过分布式推理实现联合意图，使脑机融合中的智能体进行协同决策。

## 16.9 智能科学发展路线图

通过脑科学、认知科学与人工智能领域的交叉合作,加强我国在智能科学这一交叉领域中的基础性、独创性研究,解决认知科学和信息科学发展中的重大基础理论问题,带动我国经济、社会乃至国家安全所涉及的智能信息处理关键技术的发展,为防治脑疾病和脑功能障碍、提高国民素质和健康水平等提供理论依据,并为探索脑科学中的重大基础理论问题作出贡献。

2013 年 10 月 29 日,在中国人工智能学会"创新驱动发展——大数据时代的人工智能"高峰论坛上,作者描绘了智能科学发展"路线图"[924]:2020 年,实现初级类脑计算,实现目标是计算机可以完成精准的听、说、读、写;2035 年,进入高级类脑计算阶段,计算机不但具备"高智商",还将拥有"高情商";2050 年,智能科学与纳米技术结合,发展出神经形态计算机,具有全意识,实现超脑计算。

### 16.9.1 初级类脑智能

近几年来,纳米、生物、信息和认知等当前迅猛发展的四大科学技术领域的有机结合与融合会聚成为科技界的热点,被称为 NIBC 汇聚科学技术(Converging Technologies)。这四个领域中任何技术的两两融合、三种会聚或者四者集成,都将加速科学和社会发展。脑与认知科学的进展将可能引发信息表达与处理方式新的突破,基于脑与认知科学的智能技术将引发一场信息技术的新革命。

到 2020 年,实现初级类脑计算(Elementary Brain-like Intelligence),使机器能听、说、读、写,能方便地与人沟通,突破语义处理的难关。数据的含义就是语义。数据本身没有任何意义,只有被赋予含义的数据才能够被使用,这时候数据就转化为了信息,而数据的含义就是语义。语义是对数据符号的解释,语义可以简单地看作是数据所对应的现实世界中的事物所代表的概念的含义,以及这些含义之间的关系,是数据在某个领域上的解释和逻辑表示。语义具有领域性特征,不属于任何论域的语义是不存在的。对于计算机科学来说,语义一般是指用户对于那些用来描述现实世界的计算机表示(即符号)的解释,也就是用户用来联系计算机表示和现实世界的途径。

计算机数据呈现的形态是多种多样的,目前常见的有文本、语音、图形、图像、视频、动画等。在初级阶段,机器要像人一样理解这些媒体的内容,必须突破媒体的语义理解。

### 16.9.2 高级类脑智能

到 2035 年,智能科学的目标是使机器达到高级类脑智能(Advanced Brain-like Intelligence),实现具有高智商和高情商的人造系统。智商是指数字、空间、逻辑、词汇、记忆等能力,是人们认识客观事物并运用知识解决实际问题的能力。情商是一种自我认识、了解、控制情绪的能力。情商的核心内容包括认知和管理情绪(包括自己和他人的情绪)、自我激励、正确处理人际关系三方面的能力。

### 16.9.3 超脑智能

到 2050 年，智能科学的目标是达到超脑智能（Super-brain Intelligence），实现具有意识功能的人造系统，具有高智能、高性能、低能耗、高容错、全意识等特点。

#### 1. 高智能

高智能是指由人工制造的系统所表现出来的人类水平的智能，在理解生物智能机理的基础上，对人类大脑的工作原理给出准确和可测试的计算模型，使机器能够执行需要人的智能才能完成的功能。研究智能科学建立心智模型，采用信息的观点研究人类全部精神活动，包括感觉、知觉、表象、语言、学习、记忆、思维、情感、意识等，研究人类非理性心理与理性认知融合运作的形式、过程及规律。类脑计算机实质上是一种神经计算机，它模拟人脑神经信息处理功能，通过并行分布处理和自组织方式，由大量基本处理单元相互连接而成的系统。通过大脑的结构、动力学、功能和行为的逆向工程，建立脑系统的心智模型，进而在工程上实现类心智的智能机器。智能科学将为类脑计算机的研究提供理论基础和关键技术，建立神经功能柱和集群编码模型、脑系统的心智模型，探索学习记忆、语言认知、不同的脑区协同工作、情感计算、智力进化等机制，实现人脑水平的机器智能。

#### 2. 高性能

高性能主要指运行速度。计算机的性能在 40 年内将增长 $10^8$～$10^9$ 倍，运算速度达到每秒 $10^{24}$ 次。传统的信息器件和设备系统在复杂性、成本、功耗等方面已遇到巨大障碍，基于 CMOS 的芯片技术已接近物理极限，急切期待颠覆性的新技术。另外，未来芯片要汇集计算、存储、通信等多种功能，满足多品种、短设计周期等特点，同样需要寻求新的技术路线。

硅微电子器件遵循摩尔定律和按比例缩小原则，从微米级进化到纳米级，取得了巨大成功。目前 65nm 硅 CMOS 技术已实现了大规模生产，45nm 硅 CMOS 技术开始投入生产，单芯片集成规模超过了 8 亿个晶体管。2007 年，Intel 和 IBM 同时宣布研究成功高 k 介质和金属栅技术，并应用于 45nm 硅 CMOS 技术。通过结合应变硅沟道技术和 SOI 结构，32nm 硅 CMOS 技术已试生产。Intel 和三星等公司已研究成功 10nm 以下的器件。专家们预计线宽达到 11nm 以下时，硅 CMOS 技术在速度、功耗、集成度、成本、可靠性等方面将受到一系列基本物理问题和工艺技术问题的限制。因此，在纳米级器件物理和新材料等基础研究领域不断创新、寻求突破超高速、超低功耗、亚 11nm 基础器件和集成电子系统的解决方案将成为 21 世纪世界范围最重大的科学技术问题之一。

2009 年 1 月，IBM 宣布研制出栅长为 150nm 的石墨烯晶体管，截止频率达到 26GHz。石墨烯具有极高的迁移率，是 Si 的 100 倍，饱和速度是 Si 的 6～7 倍，热导率高。适合于高速、低功耗、高集成度、低噪声和微波电路等。目前栅长 150nm 的石墨烯 MOS 晶体管的运行频率可以达到 26GHz，如果缩小到 50nm，石墨烯晶体管的频率就有望突破 1THz。预计 2020 年左右可研制成功性能优异的石墨烯材料和晶体管，并解决互联和集成等技术问题。2035 年左右可研制成功石墨烯系统芯片，并形成规模化生产。除了目前基于 CMOS 芯片的电子计算技术继续升级换代以外，量子计算、自旋电子计算、分子计算、DNA 计算和光计算等前瞻性系统技术研究正在蓬勃开展。

### 3. 低能耗

人脑运行时只消耗相当于点燃一只 20W 灯泡的能量。即使在最先进的巨型计算机上再现脑的功能,也需要一座专用的电厂。当然局部化不是唯一的差别。脑拥有一些我们还不能再现的有效率的元件。最关键的是,脑可以在大约为 100mV 的电压下工作。但是对于互补金属氧化物半导体(CMOS)逻辑电路,则需要高得多的电压(接近于 1V)才能使其正确运作,而更高的工作电压意味着在电线传送信号的过程中会用掉更多的能量。

今天的计算机执行每个运算,在电路级耗能是皮焦耳量级,在系统层是微焦耳量级,均大大高于物理学给出的理论下限。降低系统的能耗还有很大的空间。低能耗技术涉及材料、器件、系统结构、系统软件和管理模式等各个方面。从原理上创新,突破低能耗核心技术已成为今后几十年芯片和系统设计的重大挑战问题。

### 4. 高容错

容错是指一个系统在内部出现故障的情况下,仍然能够向外部环境提供正确服务的能力。容错计算的概念最早由阿维兹尼斯(A Avizienis)于 1967 年首次提出,如果一个系统的程序在出现逻辑故障的情况下仍能被正确执行,那么称这个系统是容错的。

人脑和神经网络均具有高容错的特性,部分单元失效时,仍然能够继续正确地工作。因此超脑计算系统必须具备这种高可靠的性能。

未来采用纳米级电子设备的普遍预期的制造缺陷的概率增加,文献[130]提出一种新颖的、高度容错的交叉开关体系结构。基于忆阻器交叉开关体系结构,可以使神经网络可靠实施。图 16.9 给出了单层忆阻器交叉开关矩阵。文献[130]对单层交叉开关采用 Delta 规则来学习布尔函数,呈现非常快的收敛速度。此外,在有或没有冗余的情况下,利用竞争性学习方法,模拟缺陷的影响来测量的系统结构对修复缺陷的神经元的性能。该架构能够学习布尔函数,制造缺陷率高达 13%,具有合理的冗余量。与其他技术相比,例如,级联三层冗余(Cascaded Triple Modular Redundancy)、冯·诺依曼复用和重新配置,它显示出最佳的容错性能。

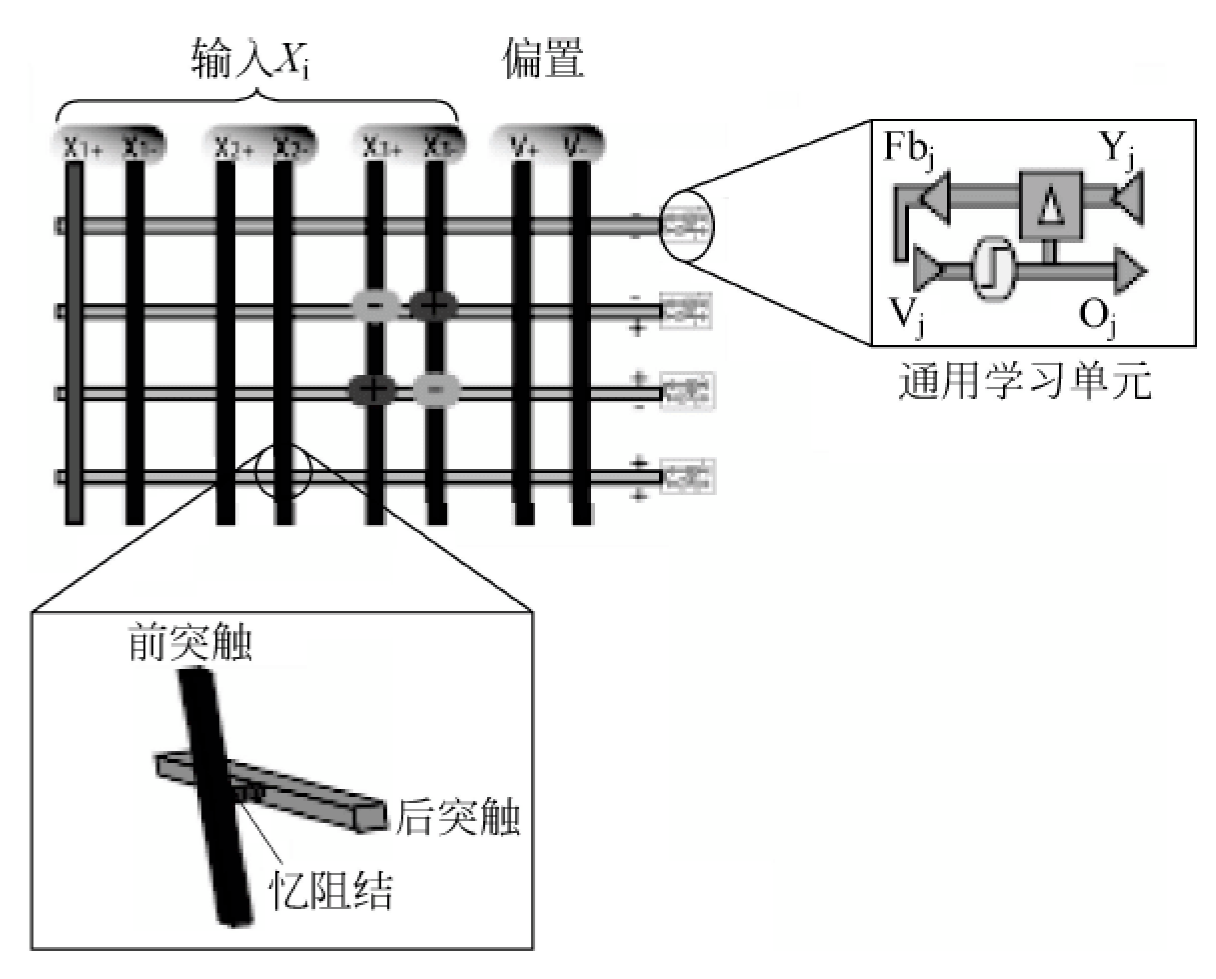



图 16.9 单层忆阻器交叉开关矩阵

### 5. 全意识

意识也许是人类大脑最大的奥秘和最高的成就之一。意识是生物体对外部世界和自身心理、生理活动等客观事物的知觉。意识的脑机制是各种层次的脑科学共同研究的对象，也是心理学研究的核心问题。人类进行意识活动的器官是脑。为了揭示意识的科学规律，建构意识的脑模型，不仅需要研究有意识的认知过程，而且需要研究无意识的认知过程，即脑的自动信息加工过程，以及两种过程在脑内的相互转化过程。同时，自我意识和情境意识也是需要重视的问题。自我意识是个体对自己存在的觉察，是自我知觉的组织系统和个人看待自身的方式，包括自我认知、自我体验、自我控制三种心理成分。情境意识是个体对不断变化的外部环境的内部表征。在复杂动态变化的社会信息环境中，情境意识是影响人们决策和绩效的关键因素。

2005 年 7 月，*Science* 杂志创刊 125 周年之际，出版了“未知之事有几何？”的专辑，提出了 125 个“有待开拓的机遇之地”问题，其中第二个问题是“意识的生物学基础是什么？”[500]。17 世纪的法国哲学家有一句名言：“我思故我在。”可以看出，意识在很长时间里都是哲学讨论的话题。现代科学认为，意识是从大脑中数以亿计的神经元的协作中涌现出来的。但是这仍然太笼统了，具体来说就是“神经元是如何产生意识的”。近年来，科学家已经找到了一些可以对这个最主观和最个人的事物进行客观研究的方法和工具，并且借助大脑损伤的病人，科学家得以一窥意识的奥秘。除了要弄清意识的具体运作方式外，还要探究它为什么存在，以及它是如何起源的。

# 参 考 文 献


[1] Aamodt A, Plaza E. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Communications, 1994, 7(1): 39-59.

[2] Abbas A K, Lichtman A H, Pober J S. Cellular and Molecular Immunology. W. B. Sarnders Company in the United Stated of American, 2003.

[3] Abbott A, Schiermeier Q. Graphene and virtual brain win billion-euro competition. Nature, 31 Jan. 2013, 493: 585-586.

[4] Abbott L F, Blum K I. Functional significance of long-term potentiation for sequence learning and prediction. Cereb. Cortex 6, 1996, 406-416.

[5] Abbott L, Dayan P. The Effect of Correlated Variability on the Accuracy of a Population Code[J]. Neural Computation, 1999, 11: 91-101.

[6] Ackley D H, Hinton G E, Sejnowski T J. A Learning Algorithm for Boltzmann Machines. Cognitive Science, 1985, 9: 147-169.

[7] Adrian E D. The Mechanism of Nervous Action. University of Pennsylvania Press, Philadelphia, 1959.

[8] Agrawal R, Imieliski T, Swami A. Mining association rules between sets of items in large database. In Proceedings of ACM SIGMOD International Conference on Management of Data(SIGMOD'93), 1993: 207-216.

[9] Alami R, Aurélie C, Vincent M, et al. Toward Human-Aware Robot Task Planning. AAAI Spring Symposium: To Boldly Go Where No Human-Robot Team Has Gone Before, 2006: 39-46.

[10] Albus J S, Bekey G A, Holland J H, et al. A proposal for a decade of the Mind Initiative [Letter]. Science 317, Sept. 7, 2007, 5843, 1321.

[11] Albus J S, MeCain H G, Lumia R. NASA/NBS Standard Reference Model for Telerobot Control System Architecture (NASREM). NBS Technical Note, 1988: 1235.

[12] Aleksander I. Impossible Minds: My Neurons, My Consciousness (Revised Edition). World Scientific Publishing Co, 2014.

[13] Aleksander I, Morton H. Aristotle's Laptop: The Discovery of our Informational Mind. World Scientific Publishing Co. Pte. Ltd, 2012.

[14] Amari S. Differential Geometrical Methods in Statistics. Springer Lecture Notes in Statistic, 28, Springer, 1985.

[15] Amari S. Information geometry of the EM and em algorithms for neural networks. Neural Networks, 1995, 8(9): 1379-1408.

[16] Amari S. Natural Gradient Works Efficiently in Learning. Neural Computation, 1998, 10: 251-276.

[17] Amoroso R L. Engineering a conscious computer. in T. Toffoli, M. Biafore, & J. Leao (Eds.), Proceedings of the 4th Workshop on Physics and Computation. Boston: N. E. Complex Systems Institute, 1996.

[18] Amoroso R L. Consciousness: A radical definition, the hard problem made easy. The Noetic Journal, 1997, 1: 1. Also in Tucson Ⅱ Abstracts, J. of Conscious Studies, 1996.

[19] Ananthanarayanan R, Esser S K, Simon H D, et al. The cat is out of the bag: Cortical simulations with 109 neurons and 1013 synapses. Gordon Bell Prize Winner. In Proceedings of the ACM/IEEE Conference on Supercomputing (Portland, OR, Nov. 14-20). ACM, New York, NY, 2009, 1-12.

[20] Ananthanarayanan R, Modha D S. Anatomy of a cortical simulator. In Proceedings of the ACM/IEEE Conference on Supercomputing, (Reno, NV, Nov. 10-16). ACM , New York, NY, 2007: 3-14.

[21] Anastasio T J. A recurrent neural network model of velocity storage in the vestibulo-ocular reflex. Advances in Neural Information Processing Systems, 1991, 3: 32-38, San Mateo, CA: Morgan Kaufmann.

[22] Anastasio T J. Modeling vestibulo-ocular reflex dynamics: From classical analysis to neural networks. in F. Eeckman, ed. Neural Systems: Analysis and Modeling, Norwell, MA: Kluwer, 1993: 407-430.

[23] Anastasio T J. Vestibulo-ocular reflex: Performance and plasticity. In M. A. Arbib, ed. The Handbook of Brain Theory and Neural Networks, Cambridge, MA: MIT Press, 1995.

[24] Anderson J A. A simple neural network generating an interactive memory. Mathematical Biosciences, 1972, 14: 197-220.

[25] Anderson J R. Language, memory, and thought. Hillsdale, NJ: Erlbaum, 1976.

[26] Anderson J R. The Architecture of Cognition. Cambridge MA, Harvard University Press, 1983.

[27] Anderson J R. The Adaptive Character of Thought. Hillsdale, NJ: Erlbaum, 1990.

[28] Anderson J R. Automaticity and the ACT* theory. American Journal of Psychology, 1992, 105: 165-180.

[29] Anderson J R. Human symbol manipulation within an integrated cognitive architecture. Cognitive Science, 2005, 29(3): 313-341.

[30] Anderson J R. How can the human mind occur in the physical universe? New York, NY: Oxford University Press, 2007.

[31] Anderson J. R. Cognitive psychology and it's implications (7th edition). New York: Worth Publishers, 2010. 安德森. 认知心理学及其启示. 7 版. 秦裕林, 等译. 北京: 人民邮电出版社, 2012.

[32] Anderson J R, Lebiere C L. The Newell test for a theory of cognition. Behavioral and Brain Science, 26, 2003: 587-637.

[33] Anderson J R, Qin Y, Sohn M-H, et al. An information-processing model of the BOLD response in symbol manipulation tasks. Psychonomic Bulletin & Review, 2003, 10: 241-261.

[34] Ashby W R. Design for a Brain. New York: Wiley, 1952.

[35] Ashby F G, Prinzmetal W, Ivry R, et al. A formal theory of feature binding in object perception. Psychological Review, 1996, 103(1): 165-192.

[36] Asperti A, Padovani L, Coen C. Sacerdoti, Schena Ⅰ. HELM and the Semantic Math-Web. In Proc. of TPHOLS 2001, Springer-Verlag, Lecture Notes in Computer Science (LNCS) Series, 2001, http://helm. cs. unibo. it/bibliography. html.

[37] Atkinson R C, Juola J F. Factors influencing the speed and accuracy of word recognition. In S. Kornblum (Ed. ), Attention and Performance Ⅳ, New York: Academic Press, 1973: 583-612.

[38] Attneave F. Some informational aspects of visual perception. Psychol. Rev. , 1954, 61: 183-93.

[39] Austin J L. How to Do Things with Words. Urmson J O, New York: OUP, 1962.

[40] Ayache N, Lustman F. Fast and reliable passive trinocular stereovision[A]. Proceedings of the International Conference in Computer Vision[C], 1987, 87: 422-427.

[41] Baars B J. A Cognitive Theory of Consciousness. New York: Cambridge University Press , 1988, Available from University Microfilm, Ann Arbor, MI, USA.

[42] Baars B J. In the Theater of Consciousness: The Workspace of the Mind, Oxford University Press, USA, 1997.

[43] Baars B J. The conscious access hypothesis: Origins and recent evidence. Trends in Cognitive Science, 2002, 6(1): 47-52.

[44] Baars B J. How brain reveals mind: neuroimaging supports the central role of conscious experience. Journal of Consciousness Studies,2003,10(9～10): 100-114.

[45] Baars B J,Banks W P,Newman J B. Essential Sources in the Scientific Study of Consciousness. A Bradford Book,2003.

[46] Baars B J,Franklin S. An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. Neural Networks,2007,20(9): 955-961.

[47] Baars B J,Franklin Stan. Consciousness is Computational: The LIDA model of global workspace theory. International Journal of Machine Consciousness,2009,1(1): 23-32.

[48] Baars B J,Gage N M (eds). Cognition,brain and consciousness (2nd Edition). Elsevier Academic Press: Burlington,MA,USA,Oxford,UK ,2010.

[49] Baars B J,Edelman D E. Consciousness,biology and quantum hypotheses. Phys Life Rev. ,2012.

[50] Bach J. A motivational system for cognitive AI. Artificial General Intelligence,LNAI 6830,2011: 232-242.

[51] Baddeley A D,Hitch G J. Working memory. In: Bower G A. The Psychology of Learning and Motivation[M]. New York: Academic Press,1974: 47-89.

[52] Baddeley A D. The episodic buffer: A new component of working memory? Trends in Cognitive Sciences,2000,4: 417-423.

[53] Baddeley A D. Is working memory still working?. American Psychologist,2001,11: 851-864.

[54] Baddeley R,Abbott L F,Booth M C,et al. Respones of neurons in primary and inferior temporal visual cortices to natural scenes. In: Proc. R. Soc. London: Ser. B,1998,264: 1775-1783.

[55] Baldi P,Brunak S. Bioinformatics: the Machine Learning Approach. Cambridge,MA: MIT Press,1998.

[56] Balduzzi D,and Tononi G. Qualia: The geometry of integrated information. PLoS Comput. Biol. 2009,5(8): 1-224.

[57] Barlow H B. Possible principles underlying the transformation of sensory messages. Sensory Communication. Edited by W A Rosenbluth,Cambridge,MA: MIT Press,1961: 217-34.

[58] Barlow H B. Single units and sensatioperceptual psychology? Perception,1972,1: 371-394.

[59] Barto A G. Reinforcement learning and adaptive critic methods. in Handbook of Intelligent Control,White D A. and Sofge D A,eds. ,New York: Van Nostrand Reinhold,1992: 469. 91

[60] Barto A G,Sutton R S,Anderson C W. Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Transactions on Systems,Man,and Cybernetics,1983,SMC-13: 834-846.

[61] Barto A G,Bradtke S J,Singh S. Learning to act using real-time dynamic programming. Artificial Intelligence,1995,72: 81-138.

[62] Bashkirov O A,Braverman E M,Muchnik I B. Potential function algorithms for pattern recognition learning machines. Automation and Remote Control,1964,25: 629-631.

[63] Bear M F,Connors B W,Paradiso M A. Neuroscience: Exploring the Brain (Third edition). Lippincott Williams & Wilkins,2006.

[64] Beaton M,Aleksander I. World-related integrated information: Enactivist and phenomenal perspectives. International Journal of Machine Consciousness,2012,4(02): 439-455.

[65] Bechara A,Damasio A R,Damasio H,et al. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition,1994,50: 7-12.

[66] Beck A T. Cognitive Therapy and Emotional Disorders. New York: International Universities. Press,1976.

[67] Beilin H,Pufall P B. Piaget's Theory: Prospects and Possibilities. New York: Psychology

Press,1992.

[68] Bell A J, Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution. Neural Computation,1995,6: 1129-1159.

[69] Berger H. Uber das elektrenkephalogramm des menschen. Arch Psychiatr Nervenkr,1929,87(1): 527-570.

[70] Bern S. Extracting Patterns and Relations from the World Wide Web. Proceedings of Web DB Workshop at EDBT'98,Valencia,1998.

[71] Berners-Lee T,Hendler J,Lassila O. The Semantic Web. Scientific American,2001.

[72] Bertsekas D P. Dynamic Programming and Optimal Control. vol. I and vol. Ⅱ,Belmont,MA: Athenas Scientific,1995.

[73] Bertsekas D P. Nonlinear Programming. Belmont,MA: Athenas Scientific,1995.

[74] Bertsekas D P,Tsitsiklis J N. Neuro-Dynamic Programming. Belmont,MA: Athenas Scientific,1996.

[75] Bhumbra G,Dyball R. Measuring spike coding in the supraoptic nucleus[J]. Journal of Physiology, 2004,555: 281-296.

[76] Bi G Q, Poo M M. Synaptic modifications in cultured hippocampal neurons: dependenceon spiketiming,synapticstrength,and postsynaptic celltype. J. Neurosci. 1998,18: 10464-10472.

[77] Biberman Y. A Context Similarity Measure. Machine Learning: ECML-94,Springer-Verlag,Berlin Heidelberg,1994.

[78] Binder J R,Frost J A,Thomas A,et al. Human brain language areas identified by functional magnetic resonance imaging. Neuroscience,1997,17(1): 353-362.

[79] Binzegger T,Douglas R J,Martin K A. A quantitative map of the circuit of cat primary visual cortex. Journal of Neuroscience,2004,24(39): 8441-8453.

[80] Birbaumer N,Schmidt R F. Biologische Psychologie. Springer-Lehrbuch,1996.

[81] Blank D, Ktunar D, Meeden L. Bringing up robot: Fundamental mechanisms for creating a self-motivated,self-organizing architecture. Cybernetics and Systems,2005,36(2): 125-150.

[82] Blumer A, Ehrenfeucht A, Haussler D et al. Classifying learnable geometric concepts with the Vapnik-Chervonenkis dimension. In Proceedings of 18th Annual ACM Symposium on Theory of Computation,Berkeley,CA,1986.

[83] Blum K I, Abbott L F. A model of spatial map formation in the hippocampus of the rat. Neural. Comput,1996,8: 85-93.

[84] Blumstein S E. Phonetic category structure and its influence on lexical processing. Proceedings of the Texas Linguistic Society,Cascadilla Press,2004.

[85] Blumstein S E, Milberg W P. Language deficits in Broca's and Wernicke's aphasia: A singular impairment. In Grodzinsky Y, Shapiro L, and Swinney D. (Eds.). Language and the Brain: Representation and Processing,Academic Press,2000.

[86] Boahen K A. A retinomorphic vision system. IEEE Micro,1996,16(5): 30-39.

[87] Boahen K A,Andreou A G. A contrast sensitive silicon retina with reciprocal synapses. Advances in Neural Information Processing Systems,San Mateo,CA: Morgan Kauftnann,1992,4: 764-772.

[88] Boahen K A,Pouliqueen P O,Andreou A G,et al. A Heterassociative memory using current-mode analog VLSI circuits. IEEE Transactions on Circuits and Systems,1989,CAS-36: 747-755.

[89] Bohm D,Hiley B J. The Undivided Universe: An ontological interpretation of quantum theory, London: Routledge,1993.

[90] Boileau A,Joyal A. La logique des topos. Journal of Symbolic Logic,1981,46(1): 6-16.

[91] Bonasso R. Integrating reaction plans and layered competences through synchronous control[A]. Proc. of IJCAI, Sydney, Australia, 1991.

[92] Booth J R, Burman D D, Meyer J R, et al. Functional anatomy of intra-and cross-modal lexical tasks, NeuroImage, 2002, 16(1): 7-22.

[93] Bower G H. Mood and memory. American Psychologist, 1981, 36(2): 129-148.

[94] Braitenberg V. Is the cerebella cortex a biological clock in the millisecond range? in The Cerebellum. Progress in Brain Research, Fox C A and Snider R S, eds., Amsterdam: Elsevier, 1967, 25: 334-346.

[95] Braitenberg V. Reading the structure of brains. Network: Computation in Neural Systems, 1990, (1): 1-12.

[96] Braitenberg V. Two views of the cerebral cortex. in Brain Theory, Palm G and Aertsen A, eds., New York: Springer-Verlag, 1986: 81-96.

[97] Braitenberg V. Vehicles: Experiments in Synthetic Psychology, Cambridge, MA: MIT Press, 1984.

[98] Braitenberg V. On the Texture of Brains, New York: Springer-Verlag, 1977.

[99] Bredeche N, Shi Zhongzhi, Zucker J. Perceptual Learning and Abstraction in Machine Learning. IEEE ICCI, 2003: 18-25.

[100] Bregman A S. Auditory Scene Analysis. Cambridge, MA: MIT Press, 1990.

[101] Breiman L. Bagging Predictors. Machine Learning, 1996, 24(2): 123-140.

[102] Brette R. Simulation of networks of spiking neurons: A review of tools and strategies. Journal of Computational Neuroscience, 2007, 23(3): 349-398.

[103] Broadbent D. Perception and Communication. London: Pergamon Press, 1958.

[104] Brooks R. A robust layered control system for a mobile robot [J]. IEEE Journal of Robotics and Antomation, 1986, 2(1): 14-23.

[105] Brooks R A. Intelligence without reasoning. In Proceedings of IJCAI'91, Sydney, 1991, 569-595.

[106] Brooks R A. Intelligent without representation. Artificial Intelligence, 1991, Vol. 47, 139-159.

[107] Broomhead D S. andLowe D. Multivariate functional interpolation and adaptive networks. Complex Systems, 1988, 2: 321-355.

[108] Brown A, Hinton G E. Products of Hidden Markov Models. In T. Jaakkola and T. Richardson eds., Proceedings of Artificial Intelligence and Statistics, Morgan Kaufmann, 3-11, 2001.

[109] Brown A, Hinton G E. Training Many Small Hidden Markov Models. in Proceedings of the Workshop on Innovation in Speech Processing, 2001.

[110] Brown R D. Example-Based Machine Translation in the Pangloss System. InProceedings of the 16th International Conference on Computational Linguistics (COLING-96), 169-174. Copenhagen, Denmark, August 5-9, 1996.

[111] Brown R D. A Modified Burrows-Wheeler Transform for Highly-Scalable Example-Based Translation. inMachine Translation: From Real Users to Research, Proceedings of the 6th Conference of the Association for Machine Translation (AMTA-2004), Washington, D. C., USA, September/October 2004: 27-36. Springer, Lecture Notes in Artificial Intelligence, Volume 3265, ISSN 0302-9743.

[112] Bruner J S. Neural mechanisms in perception. Psychological Review, 1957, 64, 340-358.

[113] Bruner J. Toward a Theory of Instruction. Cambridge, MA: Harvard University Press, 1966.

[114] Bruner J, Haste H. (Eds.). Making sense: The child's construction of the world. New York: Methuen, 1987.

[115] Bruner J. The Culture of Education. Cambridge, MA: Harvard University Press, 1996.

[116] Bryson A E, Jr., Ho Y C. Applied Optimal Control. Blaisdell, 1969, (Revised printing, 1975, Hemisphere Publishing, Washington, DC).

[117] Burges C J C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.

[118] Burges C J C. Geometry and invariances in kernel based methods, in Advance in Kernel Methods-Support Vector Learning, B. Scholkopf, C. Burge, and A. Smola, Eds., Cambridge, MA: MIT, Press, 1999, 89-116.

[119] Burke R, Kass A. Retrieving Stories for Case-Based Teaching. Case-Based Reasoning (Experiences, lessons, & Future Directions), AAAI/MIT Press, 1996, 93-110.

[120] Burks A W. An architectural theory of functional consciousness. In (N. Rescher, ed) Current Issues in Teleology. University Press of America, 1986.

[121] Burnet F M. Clonal selection and after. In: Bell G I, Perelson A S, Pimbley G H eds. Theoretical Immunology, New York: Marcel Dekker Inc., 1978, 63-85.

[122] Burnstine T H, Lesser, R P, Hart J R, et al. Characterization of the basal temporal language area in patients with left temporal lobe epilepsy. Neurology, 1990, 40(6): 966-970.

[123] Buzas P, Kovacs K, Ferecsko A S, et al. Model-based Analysis of Excitatory Lateral Connections in The Visual Cortex. JOURNAL OF COMPARATIVE NEUROLOGY, 2006, 499(6): 861-881.

[124] Buzás P, Kovács K, Ferecskó A S, et al. Model-based analysis of excitatory lateral connections in the visual cortex. Journal of Comparative Neurology, 2006, 499: 861-881.

[125] Calvin W H. How Brains Think: Evolving Intelligence, Then And Now. Basic Books, 1997.

[126] Caprotti O, Cohen A M. Draft of the OpenMath standard. The Open Math Society, 1998, http://www.nag.co.uk/projects/OpenMath/omstd/.

[127] Carlisle, D Ion P, Miner R, et al. Mathematical Markup Language (MathML) version 2.0. W3c recommendation, World Wide Web Consortium, 2001, Available at http://www.w3.org/TR/MathML2.

[128] Case R. The Mind's Staircase: Exploring the conceptual underpinnings of children's thought and knowledge, Erlbaum, 1992.

[129] Celikel T, SzostakV A, Feldman D E. Modulationof spike timing by sensory deprivation during induction of corticalmap plasticity. Nat. Neurosci, 2004, 7, 534-541.

[130] Chabi D, Klein J. Hight fault tolerance in neural crossbar. 5th International Conference on Design and Technology of Integrated Systems in Nanoscale Era (DTIS), 2010, 1-6.

[131] Chalmers D J. The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, 1996.

[132] Chalmers D J. What is a neuralcorrelate of consciousness? In T. Metzinger, ed. Neural Correlates of Consciousness: Empirical and Conceptual Issues. MIT Press, 2000.

[133] Chen L. Topological structure in visual perception. Science, 1982, 218, 699-700.

[134] Chen L. Topological structure in the perception of apparent motion. Perception, 1985, 14, 197-208.

[135] Chen L, Zhang S W, Srinivasan M. Global perception in small brains: Topological pattern recognition in honeybees, Proceedings of the National Academy of Sciences, 2003, 100: 6884-6889.

[136] Chen L. The topological approach to perceptual organization. Visual Cognition, 2005, 12: 553-637.

[137] Chen T, Du Z, Sun N, et al. DianNao: a small footprint high-throughput accelerator for ubiquitous machine learning. In Proceedings of the 19th international conference on Architectural support for programming languages and operating systems-ASPLOS' 14, pages 269-284, Salt Lake City, UT,

USA,2014.
[138] Chomsky N. Syntactic Structures. The Hague: Mouton,1957.
[139] Chomsky N. Cartesian Linguistics: A Chapter in the History of Rationalistic Thought. New York/London: Harper and Row,1966.
[140] Chomsky N. Language and Mind. New York: Harcourt,Brace and World,1968.
[141] Chomsky N. Language and Responsibility. New York: Pantheon Books,1977.
[142] Chomsky N. Rules and Representations. New York: Columbia University Press,1980.
[143] Chomsky N. Language and evolution (letter). New York Review of Books,February 1,1996.
[144] Cho S-B, Song G-B. Evolving CAM-Brain to control a mobile robot. Int. Journal of Applied Mathematics and Computation (SCI),springer-verlag,2000,111(2-3): 147-162.
[145] Churchland P. The engine of reason,the seat of the soul: a philosophical journey into the brain. Cambridge,MA,MIT Press,1995.
[146] Churchland P S. The Hornswoggle problem. Journal of Consciousness Studies,1996,3(5-6): 402-408.
[147] Churchland P S. and Sejnowski T. The computational brain. Cambridge,MA: MIT Press,1992.
[148] Ciresan D C,Meier U,Gambardella L M,et a1. Deep Big Simple Neural Nets Excel On Handwritten Digit Recognition. Neural Computation-NECO ,vol. abs/1003. 0,no. 12: 3207-3220,2010.
[149] Clark A. The dynamic challenge. Cognitive Science,1998,21(4): 461-481.
[150] Clark A. Where brain,body,and world collide. Journal of Cognitive Systems Research,1999,1: 5-17.
[151] Cohen J D,Schooler J W,eds. Scientific approaches to consciousness: 25th Carnegie Symposium on Cognition. Erlbaum,1997.
[152] Cohen M A. The construction of arbitrary stable dynamics in nonlinear neural networks. Neural Networks,1992,vol. 5,pp. 83-103.
[153] Cohen M A,Grossberg S. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems,Man,and Cybernetics,1983,vol. SMC-13,pp. 815-826.
[154] Collins A M,Quillian M R. Retrieval Time from Semantic Memory,Journal of Verbal Learning and Verbal Behavior,1969,8: 240-247.
[155] Collins A,Loftus E F. A spreading activation theory of semantic memory. Psychological Review,1975,82,407-428.
[156] Collins J J,Chow C C,Tmhoff T T. Stochastic resonance without tuning. Nature,1995,376: 236-238.
[157] Colwell B. Machine Intelligence meets neuroscience. Computer,2005,38(1): 12-15. 史俊译. 机器智能需要神经科学. 中国计算机学会通讯,2005,创刊号,68-72.
[158] Connell J. SSS: A hybrid architecture applied to robot navigation[A]. Proc. of IEEE ICRA-92[C],1992,Nice,France.
[159] Cook V,Newson M. Chomsky's Universal Grammar: An Introduction. Foreign Language Teaching and Research Press and Baickwell Publishers Ltd. ,2000.
[160] Cooper G F,Herskovits E. A Bayesian Method for the Induction of Probabilistic Networks from Data. Machine Learning,1992,9: 309-347.
[161] Cosmides L,Tooby J. The cognitive neuroscience of social reasoning. In: The new cognitive neurosciences,2nd edition,ed. M. S. Gazzaniga. MIT Press,2000.
[162] Cox M T,Ram A. Introspective multistrategy learning on the construction of learning strategies,Artificial Intelligence,1999,112: 1-55.
[163] Craik K W. The Nature of Explanation [M]. Cambridge UK: Cambridge University Press,1943.

[164] Crick F. The Astonishing Hypothesis. 1994. 弗朗西斯·克里克. 惊人的假说——灵魂的科学探索. 汪云九等译. 湖南科学技术出版社,1999.
[165] Crick F,Koch C. The problem of consciousness. Scientific American,September,1992,pp. 152-159.
[166] Crick F,Koch C. Consciousness and neuroscience. Cerebral Cortex,1998,8: 97-107.
[167] Crick F,Koch C. A Framework For Consciousness. Nature Neuroscience,2003,6: 119-126.
[168] Cui Z,Xu Y,Zeng J. Chapter 3: Social Emotional Optimization Algorithm with Random Emotional Selection Strategy. in Theory and New Applications of Swarm Intelligence, edited by Rafael Parpinelli and Heitor S. Lopes,InTech Education and Publishing,2012.
[169] Dai H P, Wright B A. Detecting signals of unexpected or uncertain durations. The Journal of the Acoustical Society of America,1995,98: 798-806.
[170] Darwin C J. Pattison H. Vowel quality changes produced by surrounding tone sequences. Journal of Experimental Psychology: Human Perception and Performance,1989,45: 333-342.
[171] Dasarathy B V. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. IEEE Computer Society Press,Los Alamitos,CA. ,1991.
[172] Dasgupta D, Attoh-Okine N. Immunity based systems: A survey. In: Proc IEEE International Conference on Systems,Man,and Cybernetics,1997,Orlando,Florida,369-374.
[173] Davidson R J, Irwin W. The functional neuroanatomy of emotion and affective style. Trends in Cognitive Sciences,1999,3: 11-21.
[174] de Castro L N, Von Zuben F J. Artificial immune system; Part 1: basic theory and application. School of Electrical and Computer Engineering,State University of Campinas,Campinas-SR,Brazil: Technical Report RT-DCA 01,1999.
[175] de Castro L N,Von Zuben F J. Clonal selection algorithm with engineering applications. In: Proc GECCO'OO,Las Vegas,Nevada,USA,36-37,2000.
[176] de Garis H, Michael K, Felix G, et al. Building an Artificial Brain Using an FPGA Based CAM-Brain Machine. Applied Mathematics and Computation Journal,North Holland,2000,111(1~4): 163-192.
[177] de Garis H,Michael K. THE CAM-BRAIN MACHINE (CBM) An FPGA Based Hardware Tool which Evolves a 1000 Neuron Net Circuit Module in Seconds and Updates a 75 Million Neuron Artificial Brain for Real Time Robot Control,Neurocomputing journal,Elsevier,2002,42: 35-68.
[178] Deadwyler S A, Hampson R E. Ensemble activity and behavior: what's the code? Science, 1995, 270: 1316-1318.
[179] Deerwester S,Dumais S T,Furnas G W,et al. Indexing by latent semantic analysis,Journal of the American Society for Information Science,1990,41.
[180] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statis. Soc. ,1977,B39: 1-22.
[181] Denett D C. Consciousness Explained. Boston: Little Brown and Company,1991.
[182] Deneve S. Bayesian Spiking Ⅰ: Inference[J]. Neural Computation,2007,20: 91-117.
[183] Deneve S. Bayesian Spiking Ⅱ: Inference[J]. Neural Computation,2007,20: 118-145.
[184] Denoeux T. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artif. Intell. 2008,172(2-3): 234-264.
[185] Denoeux T. A neural network classifier based on Dempster-Shafer theory. IEEE Transactions on System,Man and Cybernetics A,2000,30(2): 131-150.
[186] Deutsch J,Deutsch D. Attention: Some theoretical considerations. Psychological Review,1963,70: 80-90.

[187] D'haeseleer P, Forrest S, Helman P. An immunological approach to change detection algorithms: Analysis and implications. In: Proc IEEE Symposium on Security and Privacy, Las Alamitos, CA, USA, 1996, 110-119.

[188] DiGiovanna J, Mahmoudi B, Fortes J, et al. Coadaptive brain-machine interface via reinforcement learning. IEEE Trans Biomed Eng, 2009, 56(1): 54-64.

[189] Djurfeldt M. The Connection-set Algebra-A Novel Formalism for the Representation of Connectivity Structure in Neuronal Network Models. Neuroinformatic 2012, 10(3): 287-304.

[190] Djurfeldt M, Johannes H, Jochen M E, Niraj D, et al. Run-Time Interoperability Between Neuronal Network Simulators Based on the MUSIC Framework. Neuroinformatics, 2010, 8(1): 43-60.

[191] Djurfeldt M, Mikael L, Christopher J, et al. Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer. IBM Journal of Research and Development, 2008, 52(1-2): 31-42.

[192] Doi M, Suzuki K, Hashimoto S. Integrated Communicative Robot "BUGNOID"[A]. Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication[C], Berlin, Germany, 2002, 259-264.

[193] Doitsidis L, Valavanis K P, Tsourveloudis N C. Fuzzy logic based autonomous skid steering vehicle navigation[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway, NJ, USA: IEEE, 2002: 2171-2177.

[194] Domeshek E, Kolodner J, Zimring C. The Design of a Tool Kit for Case-Based Design Aids, In Artificial Intelligence in Design, Norwell, Ma. , Kluwer, 1994.

[195] Dorigo M, Gambardella L. Ant colony system: a cooperative learning approach to the traveling salesman problem [ J ]. IEEE Trans on Evolutionary Computation, 1997, 1(1): 53-66.

[196] Douglas R M, Goddard G V. Long-term potentiation of the perforant path-granule cell synapse in the rat hippocampus. Brain Res. 1975, 86, 205-215.

[197] Dreyfus H L. What Computers Can't Do. Harper&ROW, Publishers, 1979. 中译本:休伯特·德雷福斯著,宁春岩译,马希文校:计算机不能做什么.生活、读书、新知三联书店,1986.

[198] Duda R O, Hart P E. Pattern Classification and Scene Analysis, New York: Wiley, 1973.

[199] Duifhuis H, Willems L F. Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. The Journal of the Acoustical Society of America, 1982, 71: 1568-1580.

[200] Eccles J C. Brain and Conscious Experience. New York: Springer-Verlag, 1996.

[201] Eckhorn R, Reiboeck H J, Arndt M, et al. A Neural Networks for feature linking via synchronous activity: Results from cat visual cortex and from simulations, In Models of Brain Function[M], Cotterill R M J, Ed. Cambridge, Cambridge Univ. Press, 1989.

[202] Eckhorn R, Reitboeck H J, Arndt M, et al. Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. Neural Comput. , 1990, 2: 293-307.

[203] Edelman G M. Naturalizing consciousness: a theoretical framework. Proc Natl Acad Sci USA, 2003, 100(9): 5520-4.

[204] Edelman G M. Biochemistry and the sciences of recognition. J. Biol. Chem. 2004, 279: 7361-7369.

[205] Edelman G M. Learning in and from brain-based devices. Science, 2007, 318(5853): 1103-1105.

[206] Edelman G M, Tononi G. A universe of consciousness: how matter becomes imagination. New York, NY: Basic Books, 2000.

[207] Ekeberg Ö, Walle'n P, Lansner A, et al. A Computer Based Model for Realistic Simulations of Neural Networks. I: The Single Neuron and Synaptic Interaction. Biol. Cybern. 1991, 65(2): 81-90.

[208] Eliasmith C. The third contender: A critical examination of the dynamicist thory of cognition. 1996.

in Thagard P (ed) Mind Readings: Introductory Selection in Cognitive Science. MIT Press, 1998.

[209] Eliasmith C, Anderson C H. Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems. 2003, TLFeBOOK.

[210] Eliasmith C, Stewart T C, Choo X, et al. large-scale model of the functioning brain. Science, 2012, Vol. 338, 1202-1205.

[211] Ellis A W, Young A W. Human cognitive neuropsychology. Hillsdale, NJ: Erlbaum, 1988.

[212] Endoh S, Toma N, Yamada K. Immune algorithm for n-TSP. In Proc IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, USA, 1998, 3844-3849.

[213] Engel A K, Konig P, Kreiter A K, et al. Interhemispheric synchrozization of oscillatory neuronal responses in cat visual cortex. Science, 1991, 252: 1177-1179.

[214] Engel A K, Konig P, Kreiter A K, et al. Temporal coding in the visual cortex: new vistas on integration in the nervous system. Trends in Neuroscience, 1992, 15(6): 218-226.

[215] Everett H. On the Foundations of Quantum Mechanics, Ph. D. thesis, Princeton University, Department of Physics, 1957.

[216] Fairhall A, Lewen G, Bialek W, et al. Effciency and ambiguity in an adaptive neural code[J]. Nature, 2001, 412: 787-792.

[217] Farber I B, Churchland P S. Consciousness and the Neurosciences: Philosophical and Theoretical Issues. In MS Gazzaniga (ed.) The Cognitive Neurosciences (Cambridge, Massachusetts: MIT Press), 1995, 1295-1306.

[218] Farmer J D, Packard N H, Perelson A S. The immune system, adaptation, and machine learning. Physica D, 1986, 22: 187-204.

[219] Farwell L A, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials [J]. Electroencephalography and Clinical Neurophysiology, 1988, 70 (6): 510-523.

[220] Faust M E, and Gernsbacher M A. Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. Brain and Language, 1996, 53: 234-259.

[221] Ferreira F, Clifton C Jr. The independence of syntactic processing. Journal of Memory and language, 1986, 25: 348-368.

[222] Ferster D, Spruston N. Cracking the neuronal code. Science, 1995, 270: 756-757.

[223] Field D J. Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A, 1987, 4(12): 2379-2394.

[224] Fillmore C J. The Case for Case. In Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1968, 1-88.

[225] Flavell J H, Miller P H, Miller S A. 认知发展. 四版. 邓赐平, 刘明译, 上海: 华东师范大学出版社, 2002.

[226] Fodor J. A. On modularity in syntactic processing. Journal of Psycholinguistic Research, 1988, 17: 125-168.

[227] Fodor J A. The Mind Doesn't Work That Way. MIT Press, 2000.

[228] Forberg J L. Quantum Consciousness and Your Immortality. Xlibris, 1994.

[229] Fowler N, Cross S, Owens C. The ARPA-Rome Knowledge-Based Planning and Scheduling Initial. IEEE Expert, 1995, 10(1): 4-9.

[230] Fox S. Introspective multistrategy learning: Constructing a learning strategy under reasoning failure, Ph. D. Thesis, Technical Report No. GIT-CC-96-06, Georgia Institute of Technology, College

of Computing, Atlanta, GA, 1996, ftp//ftp. cc. gatch. edu/pub/ai/ram/git-cc-96-06. html.

[231] Franco L. The Use of Decoding to Analyze the Contribution to the Information of the Correlations between the Firing of Simultaneously Recorded eurons[J]. Experimental Brain Research, 2004, 155: 370-384.

[232] Franke A, Michael K. System description: MBase, an open mathematical knowledge base. In David McAllester, ed., Automated Deduction, CADE-17, no. 1831 in LNAI, 455-459. Springer Verlag, 2000.

[233] Franklin S, Graesser A C. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In Intelligent Agents Ⅲ. Berlin: Springer Verlag, 1997.

[234] Franklin S, Patterson F G Jr. The Lida Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent. Integrated Design and Process Technology, IDPT-2006, San Diego, CA, Society for Design and Process Science, 2006.

[235] Fransén E, Lansner A. A model of cortical associative memory based on a horizontal network of connected columns. Network: Comput Neural Syst. 1998, 9: 235-264.

[236] Freeman W J. Mass Action in the Nervous System, New York: Academic Press, 1975.

[237] Freeman W J. Simulation of chaotic EEG patterns with a dynamic model of the olfactory system, Biological Cybernetics, 1987, vol. 56, pp. 139-150.

[238] Freeman W J. Why neural networks don't yet fly: Inquiry into the neurodynamics of biological intelligence. IEEE International Conference on Neural Networks, 1988, vol. Ⅱ, pp. 1-7, San Diego, CA.

[239] Freeman W J. The physiology of perception. Scientific American, 1991, vol. 264 (2): 78-85.

[240] Freeman W J. Tutorial on neurobiology: From single neurons to brain chaos, International Journal of Bifurcation and Chaos in Applied Sciences and Engineering, 1992, vol. 2, 451-482.

[241] Freeman W J. Societies of Brains. Hilisdale, NJ: Lawrence Eribaum, 1995.

[242] Freeman W J. Neuraldynamics: An Exploration in Mesoscopic Brain Dynamics. London, UK: Springer-Verlag, 2000.

[243] Freitag D. Information extraction from HTML: application of a general machine learning approach. AAAI-98, 1998.

[244] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 1997, 55(1): 1119-139.

[245] Friedlander D and Franklin S. LIDA and a Theory of Mind. In Artificial General Intelligence (AGI-08), ed. Ben Goertzel and Pei Wang. Memphis, TN, USA: IOS Press, 2008.

[246] Fu X, Cai L, Liu Y, et al. A computational cognition model of perception, memory, and judgment. Science in China Series F: Information Sciences, 2013.

[247] Fu K S. Syntactic Method in Pattern Recognition. Academic Press, Ncw York, 1974.

[248] Fu K S. Syntactic Pattern Recognition and Applications, Prentice-Hall, Inc., 1982.

[249] Fujii H, Ito H. Dynamical Cell Assembly Hypothesis: Theoretical Possibility of Spatio-temporal Coding in the Cortex[J]. Neural Network, 1996, 9: 1303-1350.

[250] Fukai T. A model of cortical memory processing based on columnar organization. Biol Cybern, 1994, 70: 427-434.

[251] Fukuyama O, Suzuki T, Mabuchi K. RatCar: a vehicular neuro-robotic platform for a rat with a sustaining structure of the rat body under the vehicle. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2010.

[252] Gabbay D M. Fibring Logics, Oxford University Press, 1998.

[253] Gat E. Reliable Goal-directed reactive control for real-world autonomous mobile robots [D]. Blacksburg: Virginia Polytechnic Institute and State University, 1991.

[254] Gazzaniga M S, Ivry R, Mangun G R. Cognitive Neuroscience: The Biology of the Mind (3rd Edition). W. W. Norton, New York, 2009. 周晓林,高定国译. 认知神经科学:关于心智的生物学. 北京:中国轻工业出版社,2011.

[255] Ge Y, Jiang W X. On Consistency of Bayesian Inference with Mixtures of Logistic Regression[J]. Neural Computation, 2006, 18: 224-243.

[256] Gernsbacher M A. Language comprehension as structure building. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1990.

[257] Gers F, Garis H d, Korkin M. CoDi-1Bit: A Simplified Cellular Automata Based Neuron Model, AE97, Artificial Evolution Conference, Nimes, France, Oct. 1997.

[258] Gershon D, Sobral B W, Horton B, et al. Bioinformatics in a post-genomics age. Nature, 1997, 389: 417-422.

[259] Gerstner W, Kempter R, van Hemmen J L, et al. A neuronal learning rule for sub-millisecond temporal coding. Nature, 1996, 383: 76-81.

[260] Gesu V. A distributed architecture for autonomous navigation of robots[A]. Proc. of the 15'IEEE International workshop on computer architectures for machine perception (CMMP'00) [C]. Padova, Italy, 2000.

[261] Gibson J J. The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin, 1966.

[262] Gibson E. Principles of Perceptual Learning and Development. New York: Appleton, 1969.

[263] Gibson J J. The theory of affordances. In R. Shaw & J. Bransford (eds.), Perceiving, Acting and Knowing. Hillsdale, NJ: Erlbaum, 1977.

[264] Gibson J J. The Ecological Approach to Visual Perception. Boston: Houghton Mifflin, 1979.

[265] Gilhooly K J, Logie R H, Wetherick N E, et al. Working memory and strategies in syllogistic reasoning tasks. Memory and Cognition, 1993, 21: 115-124.

[266] Girosi F, Jones M, Poggio T. Regularization theory and neural networks architectures. Neural Computation, 1995, 7: 219-269.

[267] Globus G G. Toward a noncomputational cognitive neuroscience. Journal of Cognitive Neuroscience, 1992, 4(4): 299-310.

[268] Gloor P, Berger H. On the Electroencephalogram of Man: The Fourteen Original Peports on the Human Electroencephalogram, N. Elsevier Science Publishs, Amstedam, 1969.

[269] Gödel K. Über formal unentscheidbare Sätz der P. M. und verwandter Systeme I, Monathefte flir Math. und phvsik. 1931, 38: 173-198.

[270] Gödel K. Über eine bisher noch nicht benutzte Erweiterung des finiten standpunkten. Dialectica, 1958, 12: 280-287.

[271] Godsmark D, Brown J. A black board architecture for computational auditory scene analysis, Speech communication, 1999, 27: 353-366.

[272] Goel V, Buchel C, Frith C, & Dolan R J. Dissociation of mechanisms underlying syllogistic reasoning. Neuroimage, 2000, 12 (5): 504-14.

[273] Goertzel B. The Structure of Intelligence: A New Mathematical Model of Mind. Springer-Verlag January, Hardcover, ISBN: 3540940049, 1993.

[274] Gold E M. Language identification in the limit. Information and Control, 1967, 10: 447-74.

[275] Goldberg D E. Genetic Algorithms in Search,Optimization,and Machine Learning. Addison-Wesley, 1989.

[276] Goleman D. What makes a leader? Harvard Business Review,November-December,1998.

[277] Gordon I E. Theories of Visual Perception. New York,John Wiley and Sons,1989.

[278] Gordon A S,Hobbs J R. Formalizations of Commonsense Psychology. AI Magazine,2004,. 25(4): 49-62.

[279] Graf P,Masson M E J. (Eds.) Implicit memory: New directions in cognition,development,and neuropsychology. Hillsdale,NJ: Erlbaum,1993.

[280] Graf P,Ryan L. Transfer-appropriate processing for implicit and explicit memory. Journal of Experimental Psychology: Learning,Memory,and Cognition,1990,16,978-992.

[281] Graf P,Schacter D L. Implicit and explicit memory for new associations in normal and amnesic subjects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1985, 11: 501-518.

[282] Graves A,Fernandez S,Gomez F,et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In Proceedings of ICMS,369-376,ACM,2006.

[283] Gray C M,König P,Engel A K,et al. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature,1989,338: 334-337.

[284] Gray C M,Singer W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proceedings of the National Academy of Sciences of the United States of America,86: 1698-1702,1989.

[285] Greeno J G. Situations, mental models and generative knowledge. In: Complex information processing: The impact of Herbert A. Simon. ed. Klahr D & Kotovsky K. Erlbaum,1989.

[286] Grossberg S. Neural expectation: Cerebellar and retinal analogs of cells fired by leamable or unlearned pattern classes. Kybernetik,1972,vol. 10,49-57.

[287] Grossberg S. Adaptive pattern classification and universal recoding: Ⅰ. Parallel development and coding of neural detectors. Biological Cybernetics,1976,vol. 23,121-134.

[288] Grossberg S. Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction,illusions. Biological Cybernetics,1976,vol. 23,187-202.

[289] Grossberg S. Decision, patterns and oscillations in the dynamics of competitive systems with application to Volterra-Lotka systems,J. Theoretical Biology,1978,vol. 73: 101-130.

[290] Grossberg S. How does a brain build a cognitive code? Psychological Review,1980,vol. 87: 1-51.

[291] Grossberg S. Studies of Mind and Brain. Boston: Reidel,1982.

[292] Grossberg S. Competitive learning: From interactive activation to adaptive resonance, in Neural Networks and Natural Intelligence,S. Grossberg,ed. Cambridge,MA: MIT Press,1988.

[293] Grossberg S. Laminar cortical dynamics of visual form perception. Neural Networks 2003,16(5-6): 925-931.

[294] Grossberg S. Consciousness CLEARS the mind. Neural Networks,2007,20(9): 1040-1053.

[295] Grossberg S, Massimiliano V. Spikes, Synchrony, and Attentive Learning by Laminar Thalamocortical Circuits. Brain research,2008,1218: 278-312.

[296] Guzman A. Computer Recognition of three-dimensional objects in a visual scene. Tech. Rep. MAC-TR-59,AI Laboratory,MIT,1968.

[297] Haken H S. An Introduction. Nonequilibrium Phase-Transitions and Self-Organization in Physics, Chemistry and Biology. Springer,1977.

[298] Haken H. Synergetic computers and cognition: A top-down approach to neural nets. Berlin:

Springer-Verlag,1991.

[299] Haken H. Principle of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior, and Cognition. Springer, 1996.

[300] Ham F M, Kostanic Ⅰ. Principle of Neurocomputing for Science and Engineering. McGraw-Hill Companies, Inc. 2001.

[301] Hameroff S R, Watt R C. Information processing in microtubules. Journal of Theoretical Biology, 1982, 98 (4): 549-561.

[302] Hameroff S R. The "conscious pilot"—dendritic synchrony moves through the brain to mediate consciousness. Journal of Biological Physics, 2009, 36 (1): 71-93.

[303] Han J, Micheline K. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2000.

[304] Hansel D, Sompolinsky H. Chaos and synchrony in a model of a hypercolumn in visual cortex. J Comp Neurosci, 1997, 4: 57-79.

[305] Hansen L K, Salamon P. Neural network ensembles. IEEE Trans Pattern Analysis and Machine Intelligence, 1990, 12(10): 993-1001.

[306] Hawkins J, Sandra B. On Intelligence. Times Books, Henry Holt and Company, New York, 2004. 李蓝,刘志远译. 智能时代. 北京:中国华侨出版社,2014. 5.

[307] Hawking S W. Is Artificial Intelligence a Threat to Human Existence? BBC World Service, 2014. 12. 3. 中国日报网 chinadayli. com. cn, 2014. 12. 3.

[308] Haykin S. Neural Networks: A Comprehensive Foundation. Macmillan/IEEE Press, 1994, 1998. 叶世伟,史忠植译. 神经网络原理. 机械工业出版社,2004.

[309] HBP Sub: https://www. humanbrainproject. eu/discover/the-project/subprojects/.

[310] He J, Yao X. Drift Analysis and Average Time Complexity of Evolutionary Algorithms. Artificial Intelligence, 2001, 127(1): 57-85.

[311] Hebb D O. The organization of behavior: A neuropsychological theory. New York: Wiley, 1949.

[312] Heckerman D, Meek C, Cooper G. A Bayesian Approach to Causal Discovery. MSR-TR-97-05, February, 1997.

[313] Heckerman D. Bayesian Networks for Data Mining, Data Mining and Knowledge Discovery, 1997, 1: 79-119.

[314] Heim S, Opitz B, Muller K, et al. Phonological processing during language production: fMRI evidence for a shared. production-comprehension network, Brain Res. Cogn. Brain Res. 2003, 16: 285-296.

[315] Hieter P, Boguski M. Functional genomics: its all how you read it. Science, 1997, 278(5338): 601-602.

[316] Hinkle D, Tooney C. Applying Case-Based Reasoning to Manufacturing. AI Magazine, 1995, Springer, 65-73.

[317] Hinton G E. Deterministic Boltzmann machine learning performs steepest descent in weight-space. Neural Computation, 1989, vol. Ⅰ: 143-150.

[318] Hinton G E. Deep Learning. Invited Speaker, 29th AAAI Coference on Artificial Intelligence, Austin, Texas, USA, January 24-30, 2015.

[319] Hinton G E, Oindero S, Ten Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Comp., 2006, 18(7): 1527-1554.

[320] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[321] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines. In Rumelhart, D. E. and

McClelland,J. L. , editors, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, MIT Press, Cambridge, MA, 1986.

[322] Hochberg L R, Bacher D, Jarosiewicz B, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature, 2012, 485: 372-375.

[323] 朱志刚,徐光佑,林学訚,等.多尺度全覆盖视觉导航方法[J].机器人,1998,20(4):266-272.

[324] Hodgkin A L.(霍奇金 A)徐科,谭德培译.神经冲动的传导.北京:科学出版社,1965.

[325] Hodgkin A L, Huxley A F. A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. J. Physiol. (Lond.), 1952, 117: 500-544.

[326] Holland J H. Adaptation in Natural and Artificial Systems. University of Michigan Press, 1975.

[327] Holland J H. Properties of the bucket brigade algorithm. In Proceedings of an International Conference on Genetic Algorithms and Their Applications. Pittsburg, PA, 1985.

[328] Holland J H. Emergence-From Chaos to Order. Addison-Wesley Publishing Company, 1998.

[329] Holland J H, Holyoak K J, Nisbett R E, et al. Induction: processes of Inference, Learning, and Discovery. The MIT Press, 1986.

[330] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the USA, 1982, 9(2554).

[331] Hopfield J J. Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences of the USA, 1984, 81: 3088-3092.

[332] Hopfield J J. Pattern recognition computation using action potential timing for stimulus representation. Nature, 1995, 376: 33-36.

[333] Hopfield J J, Tank D W. Neural computation of decisions in optimization problems. Biol. Cybern, 1985, 52(14): 141-152.

[334] Hornik K M, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks, 1989, 2(2): 359-366.

[335] Huang G-B, Zhu Q-Y, Siew C-K. Extreme Learning Machine: Theory and Applications. Neurocomputing, vol. 70, 2006: 489-501.

[336] Huber D E, Shiffrin R M, Lyie K B, et al. Perception and performance in short-term word priming. Psychological Review, 2001, 108: 149-182.

[337] Humphreys G W, Lamote C, Lloyd-Jones T J. An Interactive Activation Approach to Object Processing: Effects of Structural Similarity, Name Frequencie, and Task in Normality and Pathology. Memory, 1995, 3(3/4): 535-586.

[338] Humphreys G W, Cinel C, Wolfe J, et al. Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. Vision Research, 2000, 40 (12): 1569-1596.

[339] Humphrys M. The World-Wide-Mind: Draft Proposal. Dublin City University, School of Computing, Technical Report CA-0301, 2001.

[340] Humphrys M. Distributing a Mind on the Internet: The World-Wide-Mind, ECAL-01, 2001.

[341] ImageNet: http://www.image-net.org/, 2014 Stanford Vision Lab, Stanford University.

[342] Jansen B H, Zouridakis G, Brandt M E. A neurophysiologically-based mathematical model of flash visual evoked potentials. Biol Cybern, 1993, 68: 275-283.

[343] Jansen B H, Rit V G. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern, 1995, 73: 357-366.

[344] Jerne N K. The immune system. Scientific American, 1973, 229(1): 51-60.

[345] Joachims T,Mitchell T,Freitag D,et al. Webwatcher: Machine learning and hypertext. In K. Morik and J. Herrmann, editors, GI Fachgruppentreffen Maschinelles Lernen. University of Dortmund, August,1995.

[346] Johnson J L. Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity signal invariance for images. Appl. Opt. 1994,33: 6239-6253.

[347] Johnson J L,Padgett M L. PCNN models and applications. IEEE Transactions on Neural Networks, 1999,10(3): 480-498.

[348] Johnson S. Mind Wide Open. Scribner,New York,2004.

[349] Johnson-Laird P N. Mental Models: Towards a Cognitive Science of Language [M]. Inference and Consciousness. Cambridge MA: Harvard University Press,1983.

[350] Jones R M,Laird J E,Nielsen P E,et al. Automated intelligent pilots for combat flight simulation. AI Magazine,1999,20: 27-41.

[351] Jonides J,Smith E E,Koeppe R A,et al. Spatial working memory in humans as revealed by PET. Nature,1993,363: 623-625.

[352] Judd J S. Learning in networks is hard. In Proc the 1st IEEE International Conference on Neural Networks,San Diego,CA,1987,2: 685-692.

[353] Kaelbling L P. Learning in Embedded Systems. MIT Press,Cambridge MA,1993.

[354] Kaelbling L P,Littman M L,Moore A W. Reinforcement learning: A survey. Journal of Artificial Intelligence Research,1996,4.

[355] Kaelbling L P. A special issue of machine learning on reinforcement learning. 1996,22.

[356] Kaelbling L P,Littman M L,Cassandra A R. Planning and acting in partially observable stochastic domains. Artificial Intelligence,1998,101: 99-134.

[357] Kahneman D. Attention and Effort. Englewood Cliffs,NJ: Prentice-Hall,1973.

[358] Kandel E R, Schwatz, JH, Jessell TM. Principles of Neural Science (3rd Ed.). Elsevier, NY. Amsterdam,London,Tokyo,1991.

[359] Kandel E R,Schwartz J H,Jessell T M. Principles of Neural Science (Fourth Edition). McGraw-Hill Medical,New York,2000.

[360] Kang K, Sompolinsky H. Mutual Information of Population Codes and Distance Measures in Probability Space[J]. Phys. Rev. L,2001,86: 1-4.

[361] Karp R M,Rabin M O. Efficient randomized pattern-matching algorithms. IBM J. Res. Dev. 1987,31 (2): 249-260.

[362] Kelly K. Out of Control: The New Biology of Machines,Social Systems,and the Economic World. Basic Books,1994. 东西文库译. 失控：全人类的最终命运和结局. 北京：新星出版社,2011.

[363] Kennedy J,Eberhart R. Particle swarm optimization[A]. Proe. IEEE hat. Conference On Neural Networks[C],1995,1942-1948.

[364] Kennedy J,Eberhart R C. Swarm Intelligence. Morgan Kaufmann Publishers, Inc. ,San Francisco, CA. ,2001.

[365] Kerlirzin P, Vallet F. Robustness in multilayer perceptrons. Neural Computation, 1993, vol. 5: 473-482.

[366] Khatib O. Real-time obstacle avoidance for manipulators and mobile robots [J]. The International Journal of Robotics Research,1986,5(1): 90-98.

[367] Kinsbourne M. Integrated field theroy of consciousness. In: Marcel AJ, Bisiach (eds.), Consciousness in Contemporary Science. Oxford: Clarendon Press,1988.

[368] Kintsch W. The role of knowledge in discourse comprehension: a construction-integration model. Psychological Review, 1988, 95: 163-182.

[369] Kirkpatrick S, Gelatt C D Jr, Vecchi M P. Optimization by simulated annealing. Science, 1983, 220: 671-680.

[370] Kirkpatrick S. Optimization by simulated annealing: Quantitative Studies. Journal of Statistical Physics, 1984, 34: 975-986.

[371] Kirsh D. Foundations of AI: the big issues. Artificial Intelligence, 1991, 47: 3-30.

[372] Kitano H. Challenges for massive parallelism. Proceedings of the Thirteenth International Conference on Artificial Intelligence, IJCAI-93, Chambery, Grance, Morgan Kaufmann, 81, 1993.

[373] Kleene S C. Introduction to Metamathematics. Van Nostrand Reinhold. 1957. 元数学导论上册/克林著; 莫绍揆译, 1984; 下册/克林著; 莫绍揆译. 北京: 科学出版社, 1985, 688.

[374] Knill D C, Pouget A. The Bayesian Brain: the Role of Uncertainty in Neural Coding and Computation[J]. Trends in Neuroseiences, 2004, 27(12): 712-719.

[375] Koffka K. Principles of Gestalt Psychology . Lund Humphries, London, 1935. 黎炜, 译. 格式塔心理学原理. 杭州: 浙江教育出版社, 1997.

[376] Kohlhase M. OMDoc: An open markup format for mathematical documents. Seki Report SR-00-02, Fachbereich Informatik, Universiteat des Saarlandes, 2000, http://www. mathweb. org/omdoc.

[377] Kohonen T. Self-organized formation of topologically correct feature maps. Biological Cybernetics, 1982, 43: 59-69.

[378] Kohonen T. Self-organization and associative memory. Springer Verlag, 1984.

[379] Kohonen T, Kashi S. Self-Organization of a Massive Document Collection. IEEE Transactions On Neural Networks, 2000, 11(3): 574-585.

[380] Koller D. Representation, Reasoning, Learning. Invited Talk, IJCAI'01, Seattle, 2001.

[381] Kolodner J L. An introduction to case-based reasoning. Artificial Intelligence Review, 1992, 6(1), 3-34.

[382] Kolodner J L. Case-Based Reasoning. Morgan Kaufmann, 1993.

[383] Kolodner J L. Improving Human Decision Making through Case-Based Decision Aiding, AI Magazine, 1991, 12(2): 52-68.

[384] Kolp M. Orgallizational multi-agent architectures: a mobile robot example[A]. Proc. of AAMAC [C], Bologna, Italy, 2002.

[385] Konen K W, Maure T, von der Malsburg C. A Fast dynamic link matching algorithm for invariant pattern recognition. Neural Networks, 1994, 7(6/7): 1019-1030.

[386] Kostopoulos P, Petrides M. The mid-ventrolateral prefrontal cortex: Insights into its role in memory retrieval. European Journal of Neuroscience, 2003, 17: 1489-1497.

[387] Krech D, et al. (克雷奇, 克拉奇菲尔德, 利维森等). 心理学纲要. 北京: 人民教育出版社, 1981.

[388] Krizhevsky A, Sutskever I, Hinton G. E. ImageNet Classification with Deep Convolufional Neural Networks[C]. In: Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe: 2012, 1106-1114.

[389] Krogh A, Vedelsby J. Neural network ensembles, cross validation. and active learning. In: Tesauro G, Touretzky D, Leen T eds. Advances in Neural Information Processing Systems 7, Cambridge, MA: MIT Press, 1995, 231-238.

[390] Kumaran D, Hassabis D, McClelland J L, What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. Trends in Cognitive Sciences, Vol. 20, No. 7: 512-534, 2016.

[391] Kunio Y, Jibu M. Quatum Brain Dynamics and Consciousness. John Benjamins, 1995.
[392] Kuo W J, Yeh T C, Lee J R, et al. Orthographic and phonological processing of Chinese characters: an fMRI study. Neuroimage. 2004, 21(4): 1721-1731.
[393] Kurtzberg J M. Feature analysis for symbol recognition by elastic matching. in: int'l Business Machines J. of Research and Development, 1987.
[394] Laird J, Newell A, Rosenbloom P. SOAR: An Architecture for General Intelligence. Artificial Intelligence, 1987, 33(1): 1-64.
[395] Laird J E, Kinkade K R, Mohan S, Xu J Z. Cognitive Robotics using the Soar Cognitive Architecture. CogRob 2012-The 8th International Cognitive Robotics Workshop, AAAI'12, 2012.
[396] Langley P. BACON. 1: A General Discovery System. In: Proceedings of the Second National Conference of the Canadian Society for Computational Studies of Intelligence, 1978, 173-80.
[397] Langley P W. Descriptive Discovery Processes: Experiments in Baconian Science. Ph. D. dissertation, Carngie-Mellon University, 1979.
[398] Langley P, Michalski R S. Machine Learning and Discovery, Machine Learning, 1986, 1: 363-6.
[399] Langley P, Jones R. A Computational Model of Scientific Insight. In: Sternberg, R. J. (ed.) The Nature of Creativity: Contemporary Psychological Perspectives. Cambridge University Press, 1988.
[400] Langley P, Zytkow J M. Data-Driven Approaches to Empirical Discovery, Artificial Intelligence, 1989, 40: 283-312.
[401] Langley P. Heuristics for Scientific Discovery: the Legacy of Herbert Simon. In: Langley Website, 2003.
[402] Langley Website. http://wvvw. isle. org/langlev/discoverv. html, 23 April, 2002.
[403] Langton C. (ed). Artificial Life. Vol Ⅰ, MA: Addison-Wesley, 1987.
[404] Lawvere F W. Model Theory and Topoi. A Collection of Lectures by Various Authors 978-3-540-07164-8, 1975.
[405] Lazarus R S. Cognition and motivation in emotion. American Psychologist, 1991, 46: 352-367.
[406] Leake D B. Case-based reasoning. The Knowledge Engineering Review, 1994, 9(1): 61-64.
[407] Leake D B, Plaza E. Case-Based Reasoning Research and Development. LNAI No. 1266, Springer, 1997.
[408] LeCun Y, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network [C]. In: Advances in Neural Information Processing Systems. Denver: Morgan Kaufmann, 1989, 396-404.
[409] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proc of the IEEE, 1998, 86(11): 2278-2324.
[410] LeDoux J E. The Emotional Brain. New York, Simon and Schuster, 1996.
[411] Leech G, Garside R, Bryant M. CLAWS: the tagging of the British national corpus. In Proc of 15th Int'l Conf on Cimputation Linguistics, Kyoto, Japan, 1994.
[412] Lenat D B. Cyc: A Large-Scale Investment in Knowledge Infrastructure. Communications of the ACM, 1995, 38(11): 33-38.
[413] Leng G, Brown C, Bull P, et al. Responses of magnocellular neurons to osmotic stimulation involves coactivation of excitatory and inhibitory input: An experimental and theoretical analysis[J]. Journal of Neuroscience, 2001, 21(17): 6967-6977.
[414] Lerbner A Y. Fundamental of Cybernetics. Chapman and Hall, 1972. 列别涅尔等著，刘定一译. 控制论基础. 北京：科学出版社出版，1980.

[415] Levelt W J M, Flores d'A G B. (Eds.). Studies in the perception of language. New York: Wiley, 1978.
[416] Levelt W J M. Speaking: From intention to articulation. Cambridge, MA: The MIT Press, 1989.
[417] Levelt W J M. (Ed.) Lexical access in speech production. Cambridge: Blackwell, 1993.
[418] Levelt W J M. Language. In G. Adelman, & B. H. Smith (Eds.), Encyclopedia of Neuroscience (2nd rev. and enlarged ed.) (pp. 1005-1008). Amsterdam: Elsevier Science, 1999.
[419] Levelt W J M. Models of word production. Trends in Cognitive Sciences, 1999, 3(6): 223-232.
[420] Levelt W J M. Producing spoken language: A blueprint of the speaker. In C. M. Brown, & P. Hagoort (Eds.), The neurocognition of Language (pp. 83-121). Oxford: Oxford University Press, 1999.
[421] Levelt W J M, Roelofs A P A, Meyer A S. A theory of lexical access in speech production [target paper]. Behavioral and Brain Sciences, 1999, 22 (1): 1-37.
[422] Levelt W J M, Roelofs A P A, Meyer AS. Multiple perspectives on lexical access. Reply to commentaries. Behavioral and Brain Sciences, 1999, 22 (1), 61-76.
[423] Levelt W J M, Indefrey P. The Speaking Mind/Brain: Where do spoken words come from. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), Image, Language, Brain, pp. 77-94, Cambridge, MA: MIT Press, 2001.
[424] Levelt W J M. Relations between speech production and speech perception: Some behavioral and neurological observations. In E. Dupoux (Ed.), Language, Brain and Cognitive Development: Essays in honor of Jacques Mehler, pp. 241-256, Cambridge, MA: MIT Press, 2001.
[425] Levelt W J M. Defining dyslexia. Science, 2001, 292 (5520): 1300-1301.
[426] Levelt W J M. Spoken word production: A theory of lexical access. PNAS. Proceedings of the National Academy of Sciences, 2001, 98 (23): 13464-13471.
[427] Levesque H J. A Knowledge-Level Account of Abduction. IJCAI, 1989: 1061-1067.
[428] Levesque H, Lakemeyer G. Cognitive Robotics. In Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter (Eds.), Handbook of Knowledge Representation, Elsevier, Amsterdam, 2007.
[429] Li Baodong, Shi Zhongzhi. Case retrieval based on memory network. In Shi Zhongzhi (editor): Automated Reasoning, 275-285, 1992.
[430] Li Ming and Vitanyi P. An Introduction to Kolmogorov Complexity and Its Applications. Springer Verlag, 1997.
[431] Li Xiaoli, Liu Jimin, Shi Zhongzhi. A Document Chassitie Based on Word Sementic Association. PRICAI'2000, 2000.
[432] Liaw J, Berger T. Dynamic Synapse: Harnessing the Computing Power of Synaptic Dynamics. Neurocomputing, 1999, 26-27: 199-206.
[433] Liberman A M, Cooper F S, Shankweiler D P, et al. Perception of the Speech Code. Psychol. Rev. 1967, 74: 431-461.
[434] Lieberman H, Liu H, Singh P, et al. Beating Common Sense into Interactive Applications. AI Magazine, Winter, 2004, 25(4): 63-76.
[435] Lin Long-Ji, Mitchell T M. Memory approaches to reinforcement learning in non-Markovian domains. Technical report CMU-CS-92-138, Carnegie Mellon University, 1992.
[436] Lin Tsau Young. Data Mining: Granular Computing Approach. In Proceedings of he Third Pacific-Asia Conference on Knowledge Discovery and Data Mining, April 26-28, Springer-Verlag, Lecture Notes in Artificial Intelligence series, 1999.

[437] Linsker R. From basic network principles to neural architecture (series). Proceedings of the National Academy of Sciences, USA, 1986, vol. 83: 7508-7512, 8390-8394, 8779-8783.

[438] Linsker R. Towards an organizing principle for perception: Hebbian synapses and the principle of optimal neural encoding. IBM Research Report RC12820, IBM Research, Yorktown Heights, NY., 1987.

[439] Linsker R. Towards an organizing principle for a layered perceptual network. in Neural Information Processing Systems, D. Z. Anderson, ed. pp. 485-494, New York: American Institute of Physics, 1988.

[440] Linsker R. Self-organization in a perceptual network. Computer, 1988. vol. 21: 105-117.

[441] Linsker R. An application of the principle of maximum information preservation to linear systems. Advances in Neural Information Processing Systems, 1989, vol. Ⅰ: 186-194, San Mateo, CA: Morgan Kaufmann.

[442] Linsker R. How to generate ordered maps by maximizing the mutual information between input and output signals. Neural computation, 1989, vol. Ⅰ: 402-411.

[443] Linsker R. Self-organization in a perceptual system: How network models and information theory may shed light on neural organization. Chapter 10 in Connectionist Modeling and Brain Function. The Developing Interface, S. J. Hanson and C. R. Olson, eds., pp. 351-392, Cambridge, MA: MIT Press, 1990.

[444] Linsker R. Perceptual neural organization: Some approaches based on network models and information theory. Annual Review of Neuroscience, 1990, vol. 13: 257-281.

[445] Linsker R. Designing a sensory processing system: What can be learned from principal components analysis? Proceedings of the International Joint Conference on Neural Networks, 1990, vol. 2: 291-297, Washington, DC.

[446] Linsker R. Deriving receptive fields using an optimal encoding criterion. Advances in Neural Information Processing Systems, 1993, vol. 5: 953-960, San Mateo, CA: Morgan Kaufmann.

[447] Littman M L, Sutton R S, Singh S P. Predictive Representations of State. NIPS 2001: 1555-1561.

[448] Liu S, Du Z, Tao J, et al. Cambricon: An Instruction Set Architecture for Neural Networks. ISCA 2016, Seoul, South Korea, Jun 18-22, 2016.

[449] Luger G E. Artificial Intelligence. 史忠植，张银奎，赵志崑等译. 人工智能. 北京：机械工业出版社，2003.

[450] Lundqvist M, Rehn M, Djurfeldt M, et al. Attractor dynamics in a modular network model of the neocortex. Network Computation in Neural Systems, 2006, 17: 253-276.

[451] Machens C K. Building the Human Brain. Science 338, no. 6111, November 30, 2012, 1156-1157.

[452] Mahmoudi B, DiGiovanna J, Principe J C et al. Co-Adaptive Brain-Machine Interfaces via Reinforcement Learning. in Neural Interfaces Conference, Cleveland, Ohio, 2008.

[453] Markram H. The Blue Brain Project. Nature Reviews Neuroscience, February, 2006, 7: 153-160.

[454] Markram H, Helm P J, Sakmann B. Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. J. Physiol. (Lond.) 1995, 485: 1-20.

[455] Markram H, Meier K, Lippert T, et al. Introducing the Human Brain Project. Procedia CS 7: 39-42, 2011.

[456] Markram H, Lübke J, Frotscher M et al. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science, 1997, 275: 213-215.

[457] Markram H, Sakmann B. Calcium transients in dendrites of neocortical neurons evoked by single

subthreshold excitatory postsynaptic potentials via low-voltage-activated calcium channels. Proc. Natl. Acad. Sci. USA,1994,91: 5207-5211.

[458] Markram H, Sakmann B. Action potentials propogating back into dendrites triggers changes in efficacy of single-axon synapses between layer V pyramidal cells. Soc. Neurosci. Abstr. , 1995, 21,2007.

[459] Markram H, Gerstner W, Sjöström P J. A History of Spike Timing Dependent Plasticity, Front Synaptic Neuroscience,2011,Vol. 3: 1-24.

[460] Markram H, Tsodyks M. Redistribution of synaptic efficacy: a mechanism to generate infinite synaptic input diversity from a homogeneous population of neurons with out changing absolute synaptic efficacies. J. Physiol. Paris,1996,90: 229-232.

[461] Markus B. Binaural Modeling and auditory scene analysis. IEEE ASSP Workshop,15-18,1995.

[462] Marr D. Vision: a computational investigation into the human representation and processing of visual information. W. H. Freeman,San Francisco,1982. 姚国正,刘磊,汪云九译. 视觉计算理论. 北京: 科学出版社,1988.

[463] Marslen-Wilson W D. Activation,competition,and frequency in lexical access. In: G T M Altmann ed. Cognitive Models of Speech Processing. Cambridge,MA: MIT Press,148-172,1990.

[464] Marslen-Wlison W D, Tyler L K. The temporal structure of spoken language understanding. Cognition,1980,8: 1-71.

[465] Marslen-Wilson W D, Moss H E, Halen S. Perceptual distance and competition in lexical access. Journal of Experimental Psychology: Human Perception and Performance,1996,22: 1376-1392.

[466] Maslow A H. A Theory of Human Motivation. Psychological Review,1943,50: 370-396.

[467] Matthews G,Zeidner M,Roberts R D. Emotional Intelligence: Science And Myth. MIT Press,2004.

[468] Matthews P. Relationship of firing intervals of human motor units to the trajectory of post-spike after-hyperpolarization and synaptic noise[J]. Journal of Physiology,1996,492: 597-628.

[469] Mayer J,Wildgruber D,Riecker A,et al. Prosody Production and Perception: Converging Evidence from fMRI Studies. Proceedings of Speech Prosody 2002, Aix-en-Provence, France, April, 11-13, 2002,487-490.

[470] Mayer J,Ackermann H,Dogil G, et al. Syllable Retrieval vs. Online Assembly: fMRI Examination of the Syllabary. Proceedings of the XVth International Congress of Phonetic Sciences (ICPhS), Barcelona,2541-2544,2003.

[471] Mayer-Schönberge V,Cukier K. Big Data: A Revolution That Will Transform How We Live,Work, and Think. Houghton Mifflin Harcourt,2012. 盛扬燕,周涛,译. 大数据时代. 杭州: 浙江人民出版社,2013.

[472] Mayeux R,Kandel E R. Disorders of Language: The Aphasias. In: Kandel ER,Schwartz JH and Jessell TM (eds). Principles of Neural Science,3rd ed. Elsevier,1991,pp 840-851.

[473] McCallum R. A. Overcoming incomplete perception with utile distinction memory. In The Proceedings of the Tenth International Machine Learning Conference, Morgan Kaufmann Publishers,Inc. ,1993.

[474] McCarthy J. Circumscription-a form of non-monotonic reasoning. Artificial Intelligence,1980,13(1-2): 27-39.

[475] McCarthy J. Applications of circumscription to formalizing commonsense knowledge. Artificial Intelligence,1986,28: 89-116.

[476] McClelland J L. Stochastic interactive processes and the effect of context on perception. Cognitive

Psychology,1991,23: 1-44.
[477] McClelland J L. The organization of memory: A Parallel Distributed Processing perspective. Revue Neurologique (Paris),1994,150,8-9,570-579.
[478] McClelland J L. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 102,419-457,1995.
[479] McClelland J L,Elman J L. The TRACE Model of Speech Perception. Cognitive Psychology,1986,18: 1-86.
[480] McClelland J L,Rumelhart D E. Parallel distributed processing: Explorations in Parallel Distributed Processing: MIT Press,1986.
[481] McClelland J L,Mirman D,Holt L L. Are there interactive processes in speech perception? Trends in Cognitive Sciences,2006,10(8): 363-369.
[482] McClelland J L,Rumelhart D E. Explorations in Parallel Distributed Processing: MIT Press,1988.
[483] McCulloch W S, Pitts W. A logic calculus of the ideas immanent in nervous activity. Bulletin of Mathematical biophysics,1943,Vol. 5,115-133.
[484] McCulloch W S. Embodiments of Mind. MIT Press,Cambridge,MA,1988.
[485] McDermort K B. Implicit memory. In Kazdin A E. (Ed.) The encyclopedia of psychology. New York: American Psychological Association and Oxford University Press,2000.
[486] McDermott K B. Explicit and implicit memory. Encyclopedia of the human brain,2002,vol. 2: 773-781. V. S. Ramachandran (Ed.),Academic Press.
[487] McMillan G R, et al. Direct brain interface utilizing self-regulation of steady—atate visual evoked response(ssver),Proceedings of RESNA,693-695,1995.
[488] Mckoon G,Ratcliff R. Memory-based language processing: psycholinguistic Research in the 1990s. Annual Review of psychology,1998,49: 25-42.
[489] Mead C A. Neuromorphic electronic systems. Proceedings of the Institute of Electrical and Electronics Engineers,1990,vol. 78: 1629-1636.
[490] Mead C A. Analog VLSI and Neural Systems. Reading,MA: Addison-Wesley,1989.
[491] Mead C A,Mahowald M A. A silicon model of early visual processing. Neural Networks,1988,vol. 1: 91-97.
[492] Mead C A, Arreguit X,Lazzaro J. Analog VLSI model of binaural hearing. IEEE Transactions on Neural Networks,1991,vol. 2,pp. 232-236.
[493] Meersman R A, Shi Zhongzhi, Kung Chen-Ho. (Eds.) Artificial Intelligence in Databases and Information Systems (DS-3). North-Holland Publishers,1990.
[494] Melis E, Buedenbender J., Andres E, et al. Knowledge representation and management in ActiveMath. International Journal on Artificial Intelligence and Mathematics, Special Issue on Management of Mathematical Knowledge,2003,38(1-3): 47-64.
[495] Merolla P A,Arthur J V,Alvarez-Icaza R,et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science,8 Aug. 2014,Vol. 345,Issue 6197: 668-672.
[496] Metropolis N,Rosenbluth A,Rosenbluth M,et al. Equations of state calculations by fast computing machines. Journal of Chemical Physics,1953,vol. 21: 1087-1092.
[497] Meyer D E. On the representation and retrieval of stored semantic information. Cognitive Psychology,1970,1: 242-300.
[498] Meyer D E. Dual memory-search of related and unrelated semantic categories. Paper presented at the

meeting of the Eastern Psychological Association, New York, April, 1971.

[499] Meyers J L, O'Brien E J, Albrecht J E, et al. Maintaining global coherence during reading. Journal of Experimental Psychology, 1994, 20: 876-886.

[500] Miller G. What Is the Biological Basis of Consciousness? Science, 1 JULY, 2005, Vol. 309, pp. 79.

[501] Miller G A, Beckwith R, Fellbaum C, et al. Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography, 3(4): 235-244, 1990.

[502] Miller E K, Cohen J D. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience, 2001, 24: 167-202.

[503] Minsky M. A Framework for Representing Knowledge. The Psychology of Computer Vision, P. H. Winston (ed.), McGraw-Hill, 1975.

[504] Minsky M. The Society of Mind. New York, Simon & Schuster, 1985. 马文·明斯基. 心智社会. 任楠译. 北京: 机械工业出版社, 2016.

[505] Minsky M. Semantic Information Processing. Cambridge, MA.: MIT Press, Hall, R. P., 1989.

[506] Minsky M. Machinery of Consciousness. Proceedings, National Research Council of Canada, 75th Anniversary Symposium on Science in Society, June, 1991.

[507] Minsky M. The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Psychology & Psychotherapy, Science, Barnes & Noble. com, 2006. 马文·明斯基著. 情感机器. 王文革, 程玉婷, 李小刚译. 杭州: 浙江人民出版社, 2015.

[508] Minsky M, Papert S. Perceptrons. MIT Press, 1969.

[509] Misiti M, Misiti Y, Oppenheim G, et al. Wavelef Toolbook for use with Matlab. Natick MA: The Math Works Inc, 1997.

[510] Mitchell T. Machine Learning. McGraw Hill, 1997.

[511] Miyake A, Shan P. Models of working memory: mechanisms of active maintenance and executive control. New York: Cambridge University Press, 1999.

[512] Modha D S, Ananthanarayanan R, Esser S K., et al. Cognitive computing. Commun. ACM, 2011, 54 (8): 62-71. 史忠植译. 认知计算. 中国计算机学会通讯, 2012, 8(8): 76-84.

[513] Modha D S, Singh R. Network architecture of the long-distance pathways in the macaque brain. Proceedings of the National Academy of Sciences of the USA 107, 30, June, 2010, 13485-13490.

[514] Mooney R J. Inductive logic programming for natural language processing. Inductive Logic Programming, 1997, 1314: 3-22.

[515] Moravec H(汉斯·莫拉维克). Robot: mere machine to transcendent mind. Oxford University Press, November, 1998. 马小军, 时培涛译. 机器人. 上海: 上海科学技术出版社, 2001.

[516] Moravec H P. Visual mapping by a robot rover[A]. Proceedings of the 6th International Joint Conference on Artificial Intelligence[C], Tokyo, Japan, 1979, 599-601.

[517] Mountcastle V B. Modality and topographic properties of single neurons of cat's somatic sensory cortex. J Neurophysiol, 1957, 20: 408-434.

[518] Mountcastle V B. The columnar organization of the neocortex. Brain, 1997, 120: 701-722.

[519] Mummery C J, Patterson K, Hodges J R, et al. Functional neuroanatomy of the semantic system: divisible by what? J Cogn Neurosci, 1998, 10(6): 766-777.

[520] Murphy S T, Zajonc R B. Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. Journal of Personality & Social Psychology, 1993, 64(5): 723-739.

[521] Nagel H. Steps toward a Cognitive Vision System. AI Magazine, 2004, 25(2): Summer, 31-50.

[522] Nakano H, Blumstein S E. Deficits in thematic integration processes in Broca's and Wernicke's

aphasia. Brain and Language, 2004, 88: 96-107.

[523] Nathans D, Smith H O. Restriction Endonucleases in the Analysis and Restructuring of DNA Molecules. Ann. Rev. Biochem. , 1975, 44: 273-29.

[524] Neal R M. Bayesian Learning for Neural Networks, Ph. D. Thesis, University of Toronto, Canada, 1995.

[525] Neal R M. Bayesian learning via stochastic dynamics. Advances in Neural Information. Processing Systems, 1993. vol. 5: 475-482, San Mateo, CA: Morgan Kaufmann.

[526] Neal R M. Connectionist learning of belief networks. Artificial Intelligence, 1992, Vol. 56, pp. 71-113.

[527] Neisser U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

[528] Newell A. Productions systems: models of control structures. Pittsburgh, Pa: Carnegie Mellon University, Department of Computer Science, 1973.

[529] Newell A. Physical symbol systems. Cognitive Science, 1980, 4: 135-183.

[530] Newell A. Physical symbol systems. in Perspectives on cognitive science. Norman, D. A. , ed. Hillsdale, N. J. : Lawrence Erlbaum Associates, 1981.

[531] Newell A. Unified Theories of Cognition. Cambridge, Mass. : Harvard University Press, 1990.

[532] Newell A, Simon H A. The logic theory machine: A complex information processing system. IRE Trans. Inf. Theory, 1956, IT-2: 61-79.

[533] Newell A, Simon H A. Human Problem Solving. Englewood Cliffs, N. J. : Prentice-Hall, 1972.

[534] Newell A, Simon H A. Computer science as empirical inquiry: Symbols and search. Communications of the Association for Computing Machinery, 1976, 19(3): 113-126 (1975 ACM Turing Award Lecture. ).

[535] Newell A, Simon H A. Symbol manipulation. In Encyclopedia of computer science A Ralston and C L Meek (Eds. ) New York: Petrocelli/Charter, 1384-1389, 1976.

[536] Newell A, Laird J E, Rosenbloom P S. SOAR: An Architecture for General Intelligence. Artificial Intelligence, 1987, 33(1): 1-64.

[537] Newell A, Rosenbloom P S, Laird J E. Knowledge Level Learning In Soar. AAAI 1987: 499-504, 1987.

[538] Newell A, Rosenbloom P S, Laird J E. (eds. ) The Soar papers: research on integrated intelligence. Cambridge, Mass. MIT Press, 1993.

[539] Nicholls J G. (J G 尼克尔斯等)杨雄里等译. 神经生物学——从神经元到脑. 北京: 科学出版社, 2003.

[540] Nilsson N. Learning Machines. McGraw-Hill, New York, 1965.

[541] Nicolis G, Prigogine I. Exploration of Complexity. 1986. 罗久里等译. 探索复杂性. 成都: 四川教育出版社, 1992.

[542] Nirenberg S, Carcieri S M, Jacobs A L, et al. Retinal ganglion cells act largely as independent encoders. Nature, 2001, 411: 698-701.

[543] Norman D A. Twelve issues for cognitive science. in Perspectives on cognitive science. Norman, D A, ed. Hillsdale, N. J. : Lawrence Erlbaum Associates, 1981.

[544] Norman D A. What is cognitive science? in Perspectives on cognitive science. Norman, D A, ed. Hillsdale, NJ. : Lawrence Erlbaum Associates, 1981.

[545] Norman D A. Emotional Design: Why We Love (or Hate) Everyday Thing. Apress L. P. 2003. 情感化设计. 付秋芳, 程进三译, 严正, 傅小兰审校, 北京: 电子工业出版社, 2005.

[546] O'Doherty J E, Lebedev M A, Ifft P J, et al. Active tactile exploration using a brain-machine-brain

interface. Nature,2011,479(7372):228-231.
[547] Painkras E, Plana L A. SpiNNaker: A 1-W 18-Core System-on-Chip for Massively-Parallel Neural Network Simulation. IEEE JOURNAL OF SOLID-STATE CIRCUITS,2013,48(8):1943-1953.
[548] Paninski L. Estimation of Entropy and Mutual Information[J]. Neural Computation,2003 15:1191-1253.
[549] Pavlov I P. http://www.iemrams.spb.ru:8101/english/pavlov.htm.
[550] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers,1992.
[551] Pearl J. Causality: Models,Reasoning,and Inference. New York: Cambridge University Press 2000.
[552] Penrose R. The Emperor's New Mind. Oxford University Press,1989.
[553] Penrose R. Shadows of the Mind. Oxford Press,London,1994.
[554] Penrose R, Hameroff S R. Conscious Events as Orchestrated Space-Time Selections. Journal of Consciousness Studies,1996,3 (1):36-53.
[555] Penrose R, Hameroff Stuart. Consciousness in the Universe: Neuroscience, Quantum Space-Time Geometry and Orch OR Theory. Journal of Cosmology,2011,Vol. 14.
[556] Pentland A P. Perceptual organization and the representation. of natural form. Artificial Intelligence, 1986,28(2):293-331.
[557] Perelson A. Immune network theory. Immunological Review,1989,110:5-36.
[558] Perkowitz M. Etzioni O. Category translation: Learning to understand information on the Internet. Proceedings of IJCAI-95,1995,Montreal,Morgan Kaufmann.
[559] Perkowitz M, Etzioni O. Adaptive Web Sites: Automatically Synthesizing Web Pages. AAAI98,1998.
[560] Perus M. Neuro-Quantum parallelism in brain-mind and computers, Informatica, 1996 vol. 20: 173-183.
[561] Perus M. Mind: neural computing plus quantum consciousness. in: M. Gams, M. Paprzycki, X. Wu (Eds.). MIND VERSUS COMPUTER. IOS Press, Amsterdam,1997, pp. 156-170.
[562] Perus M. Visual conscious experience. Ch. 5 in R. Amoroso (Ed.). The Complementarity of Mind and Body. Nova Sci. Publ. ,New York,2010,pp. 75-113.
[563] Petrides M. Functional organization of the human frontal cortex for mnemonic processing. Ann. N. Y Acad. Sci. ,1995,769:85-96.
[564] Petrides M, Alivisatos B, Evans A C. et al. Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. Proc. Nut1 Acad. Sci. USA,1993,90:873-877.
[565] Pfurtscheller G,et al. Prediction of the side of hand movements from single trial multichannel EEG data using neural networks,Electroenceph Ctin Neurophysiol,1992,82:313-315.
[566] Piaget J. The psychology of intelligence. New York: Routledge,1963,2001.
[567] Piaget J. Genetic epistemology. New York: W. W. Norton & Company,1970.
[568] Piaget J. 发生认识论原理. 王宪钿等译. 北京:商务印书馆,1972.
[569] Piaget J. Intellectual evolution from adolescence to adulthood. Human Development,1972,15(1):1-12.
[570] Piaget J. Memory and intelligence. New York: BasicBooks,1973.
[571] Piaget J. Adaptation and intelligence: Organic selection and phenocopy. Chicago: University of Chicago Press,1974,1980.
[572] Piaget J. Le Structuralism. 倪连生,王琳译. 结构主义. 北京:商务印书馆,1979.
[573] Piaget J, Henriques G, Ascher E, et al. 态射与范畴:比较与转换. 刘明波,张兵,孙志凤译. 上海:华东师范大学出版社,2005.

[574] Piaggio M. HEIR-A non hierarchical architecture for intelligent robots[A]. ATAL'98[C], Paris, France, 1998.
[575] Picard R W. Affective Computing. MIT Press, London, England, 1997.
[576] Pink D H. 全新思维. 林娜译. 北京：北京师范大学出版社, 2006.
[577] Pinker S. The language instinct: How the mind creates language. New Your: Morrow, 1994.
[578] Poggio T. A theory of how the brain might work. Cold Spring Harbor Symposium on Quantitative Biology, 1990, vol. 5, pp. 899-910.
[579] Poggio T, Beymer D. Learning to see. IEEE Spectrum, 1996, 33(5): 60-69.
[580] Poggio T, Edelman S. A network that learns to recognize three-dimensional objects. Nature, 1990, vol. 343, pp. 263-266.
[581] Poggio T, Fahle M, Edelman S. Fast perceptual learning in visual hyperacuity. Science, 1992, 256: 1018-1021.
[582] Poggio T. Girosi F. A Theory of Networks for Approximation and Learning. Technical Report 1140, MIT AI Lab. 1989.
[583] Poggio T, Girosi F. Networks for approximation and learning. Proceedings of the IEEE, 1990, vol. 78: 1481-1497.
[584] Poggio T, Girosi F. Reguiarization algorithms for learning that are equivalent to multilayer networks. Science, 1990, vol. 247: 978-982.
[585] Poggio T, Koch C. Ill-posed problems in early vision: From computational theory to analogue networks. Proceedings of the Royal Society of London, Series B, 1985, vol. 226: 303-323.
[586] Poggio T, Torre V, Koch C. Computational vision and regularization theory. Nature, 1985, vol. 317: 314-319.
[587] Posner M I. Chronometric explorations of mind. Hillsdale, NJ: Erlbuum, 1978.
[588] Posner M I. Attention: the mechanism of consciousness. Proc. National Acad of Sciences, USA., 1994, 91(16): 7398-7402.
[589] Posner M I. Imaging a science of mind. Trends in Cognitive Sciences, 2003, 7(10): 450-453.
[590] Posner M I. The achievements of brain imaging: Past and present. To appear in N. Kanwisher & J. Duncan (Eds.), Attention and Performance XX, Oxford University Press pp. 505-528, 2004.
[591] Posner M I. (ed). Cognitive Neuroscience of Attention. New York: Guilford, 2004.
[592] Posner M I, DiGirolamo G J. Attention in cognitive neuroscience: an overview. In M. S. Gazzaniga (ed.) The New Cognitive Neurosciences Second Edition 621-632, 2000.
[593] Posner M I, Raichle M E. Images of mind. New York: Freeman, 1994.
[594] Pouget A, et al. Statistically Efificient Estimation Using Population Codes[J]. Neural Computation, 1998, 10: 373-401.
[595] Pouget A, Dayan P, Zemel R. Inference and Computation with Population Codes[J]. Annu Rev Neurosei. 2003, 26: 381-410.
[596] Powell M J D. The Theory of Radial Basis Function Approximation. Cambridge University Numerical Analysis Report, 1990.
[597] Powell M J D. Radial basis functions for multivariate interpolation. in Algorithms for Approximation, eds. Mason, J. M. and Cox, M., 143-167, 1985.
[598] Power M, Dalgleish T. Cognition and Emotion: From Order to Disorder. Hove: Psychology Press, 1997.
[599] Prigogine I, Kondepudi D. Modern Thermodynamics: From Heat Engines to Dissipative Structures.

John Wiley & Sons,Chichester,1998.

[600] Prigogine I,Nicolis G. Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations. Wiley & Sons,New York,1977.

[601] Prinzmetal M,Diedrichsen S,Ivry R B. Illusory conjunctions are alive and well: A reply to Donk. Journal of Experimental Psychology: Human Perception and Performance,2001,27(3): 538-541.

[602] Prinzmetal M,Ivry R B,Beck D,et al. A measurement theory of illusory conjunctions. Journal of Experimental Psychology: Human Perception and Performance,2002,28(2): 251-269.

[603] Pylyshyn Z. What the mind's eye tells the mind's brain. Psychological Bulletin,1973,80: 1-24.

[604] Pylyshyn Z. Imagery theory: Not mysterious-just wrong. Behavoral and Brain Sciences,1979,2: 561-563.

[605] Pylyshyn Z. Return of the mental image: Are there pictures in. the brain? Trends in Cognitive Sciences,2003,7: 113-118.

[606] Quillian M R. Semantic Memory, in Marvin Minsky (ed.) Semantic Information Processing, Cambridge,MA: MIT Press,1968,227-270.

[607] Quillian M R. The Teachable Language Comprehender: A Simulation Program and Theory of Language,Communications of the Association for Computing Machinery,1969,12(8): 459-476.

[608] Quinlan J R. Discovering rules from large collections of examples: A case study. in Michie D. (editor). Expert Systems in the Micro Electronic Age,Edinburgh University Press,1979.

[609] Quinlan J R. Learning efficient classification procedures and their application to chess end-games. in R S Michalski, J G Carbonell and T M Mitchell (editors), Machine Learning: An Artificial Intelligence Approach,Tioga,1983.

[610] Quinlan J R. Induction of decision trees. Machine Learning,1986,1(1): 81-106.

[611] Quinlan J R. The effect of noise on concept learning. in R S Michalski, J G Carbonell and T M Mitchell (editors), Machine Learning: An Artificial Intelligence Approach, Morgan Kaufmann,1986.

[612] Quinlan J R. Simplifying Decision Trees, Internat. Journal of Man-Machine Studies,1987, Vol. 27: 221-234.

[613] Quinlan J R. Generating production rules from decision trees. in Proceedings of IJCAI-87, Milan, Italy,1987.

[614] Quinlan J R. Simplifying Decision Trees, Internat. Journal of Man-Machine Studies,1987, Vol. 27: 221-234.

[615] Quinlan J R. An empirical comparison of genetic and decision-tree classifiers. In Proceedings of ICML-88,San Mateo,CA. ,1988.

[616] Quinlan J R. C4. 5: Programs for Machine Learning. Morgan Kaufmann,1992.

[617] Quillian M R. Semantic memory. In M. Minsky, editor, Semantic Information Processing, pages 216-270. The MIT Press,1968.

[618] Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE,1989,77(2): 257-286.

[619] Radford A. Syntax: A Minimalist Introduction. Foreign Language Teaching and Research Press and Cambridge University Press,2000.

[620] Ram A, Arkin R, Boone G. Using genetic algorithms to learn reactive control parameters for autonomous robotic navigation [J]. Adaptive Behavior,1994,2(3): 277-304.

[621] Reeke G, Coop A. Estimating the temporal interval entropy of neuronal discharge[J]. Neural Computation,2004,16: 941-970.

[622] Reiter R. Knowledge in action: logical foundations for specifying and implementing dynamical systems. MIT Press, Cambridge, Mass, 2001.
[623] Richmond B J, Optican L M, Podell M. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. J Neurophys, 1987, 57(1): 132-178.
[624] Riehle A, Grun S, Diesmann M, et al. Spike synchronization and rate modulation differentially involved in motor cortical function. Science, 1997, 278: 1950-1953.
[625] Rieke F, et a1. Spikes: Exploring the Neural Code[M]. Cambridge: MIT Press, 1997.
[626] Robertson S S, Cohen A H, Mayer-Kess R G. Behavioural Chaos: Beyond the Metaphor, in SMITH, L. B. & THELEN, E. (Eds), A dynamic systems approach to development: Applications (Cambridge, MIT Press), 1993 pp. 120-150.
[627] Robin N, Holyoak J. Relational Complexity and the Functions of Prefrontal Cortex. In The Cognitive Neurosciences, VIII THOUGHT AND IMAGERY, M. S. Gazzaniga, ed. London: The MIT Press, 1995.
[628] Roco M C, Bainbridge WS. (Eds.) Converging Technologies for Improving Human Performance. NSF/DOC-Sponsored Report, 2002.
[629] Rolls E T, et al. Information Encoding in the Inferior Temporal Visual Cortex: Contribution of the Firing Rates and the Correlations between the Firing of Neurons[J]. Biological. Cybernetics, 2004, 90: 19-32.
[630] Rooney B. The social robot architecture: Towards sociality in a real world domain[A]. Towards intelligent mobile robots 99[C], Bristol, 1999.
[631] Rosenblatt F. The Perceptron: A probabilistic model for information storage and organization in the brain. Phychological Review, 1958, 65: 386-408.
[632] Rosenblatt F. Principles of Neurodynamics. Spartan Books, 1962.
[633] Rosenblatt J. DAMN: A distributed architecture for mobile navigation[D]. Robotics Institute, 1997.
[634] Roy D. Learning from Sights and Sounds: A Computational Model. Ph. D. Thesis, MIT Media Laboratory, 1999.
[635] Rumelhart D E, McClelland J L, et al. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I, Cambridge, MA: MIT Press, 1986.
[636] Rumelhart D E, Zipser D. Feature discovery by competitive learning. Cognitive Science, 1985, 9: 75-112.
[637] Rumelhart D E, Hinton G E, Williams R J. Learning representations of back-propagation errors. Nature (London), 1986, 323: 533-536.
[638] Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by error propagation. in D. E. Rumelhart and J. L. McCleland, eds. , vol I, Chapter 8, Cambridge, MA: MIT Press, 1986.
[639] Russell B, Whitehead A. Principia Mathematica. 3 volumes, 1910-1913.
[640] Salovey P, Mayer J D. Emotional Intelligence. Imagination, Cognition and Personality, 1990, 9: 185-211.
[641] Saridis G. Toward the realization of Intelligent Controls. Proc. of the IEEE, 1979, 67(4): 1115-1133.
[642] Samuel A L. Some studies in machine learning using the game of checkers. in E A Feigenbaum and J Feldman (editors), Computers and Thought, McGraw-Hill, 1963.
[643] Sato M, Sato Y, Jain L C. Fuzzy Clustering Models and Applications. Springer-Verlag Company, 1997.
[644] Saul L K, Jakkolla T, Jordan M I. Mean field theory for sigmoid belief networks. Journal of Artificial Intelligence Research, 1996, vol. 4: 61-76.
[645] Saul L K, Jordan M I. Exploiting tractable substructures in intractable networks. Advances in Neural

Information Processing Systems, vol. 8, pp. 486-492, Cambridge, MA: MIT Press, 1996.
[646] Saul L K, Jordan M I. Boltzmann chains and hidden Markov models. Advances in Neural Information Processing Systems, 1995, vol. 7: 435-442.
[647] Schaffer C. Overfitting Avoidance as Bias. in: ML, 1993, vol. 10: 153-178.
[648] Schank R C. Conceptual Dependency: {A} Theory of Natural Language Understanding, Cognitive Psychology, 1972, (3)4: 532-631.
[649] Schank R C, Abelson R P. Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures (Chap. 1-3), L. Erlbaum, Hillsdale, NJ, 1977.
[650] Schank R C. Dynamic Memory: A Theory of Learning in Computers and People. New York: Cambridge University Press, 1982.
[651] Schacter D L. Implicit memory: history and current status. Journal of Experimental Psychology: L. M, & C. 1987, 13(3): 501-518.
[652] Schacter D L, McAndrews M P, Moscovitch M. Access to consciousness: dissociations between Imphcit knowledge in neuropsychological syndromes. In: WeiskantzL (ed). Thought without Language. Oxford: Oxford University Press. 242-278, 1988.
[653] Schank R C. Conceptual Information Processing. New York: Elsevier, 1975.
[654] Schank R C. Language and Memory. Cognitive Science, 1980, vol. 4: 243-284.
[655] Schank R C. Dynamic Memory: A Theory of Reminding and Learning in Computers and People. Cambridge University Press, 1982.
[656] Schemmel J, Brüderle D, Meier K, et al. A wafer-scale neuromorphic hardware system for large-scale neural modeling. ISCAS 2010: 1947-1950.
[657] Schmidhuber J. Deep Learning in Neural Networks: An Overview. http://www.idsia.ch/-juergen/deep-learning-overview.html, 2014.
[658] Schnitzer M J, Meister M. Multineuronal Firing Patterns in the Signal from Eye to Brain. Neuron, 2003, 37: 99-115.
[659] Schreiber G. Knowledge Engineering and Management. 2000. 史忠植,梁永全,吴斌等译. 知识工程和知识管理. 北京: 机械工业出版社, 2003.
[660] Schulze-Kremer S. Molecular Bioinformatics: Algorithms and Applications. Walter de Gruyter. Berlin, New York, 1996.
[661] Schumpeter J A. The Theory of Economic Development. Cambridge, MA: Harvard University Press, 1934. 熊彼特, 经济发展理论. 北京: 商务印书馆, 中译本, 1990.
[662] Schunpeter J A. Business Cycles: A Theoreitcal, Histoircal, and Statistical Analysis of the Capital Process. New York: McGraw-Hill, 1939.
[663] Schumpeter J A . Capitalism, Socialism, and Democracy. New York, 1942, 熊彼特. 社会主义, 资本主义与民主. 北京: 商务印书馆, 中译本, 1992.
[664] Schuster H G, Wagner P. 1: A model for neuronal oscillations in the visual cortex. 2: Phase description of the feature dependent synchronization. Biol Cybern, 1990, 64: 77-85.
[665] Schwaighofer A, Volker T. The Bayesian committee support vector machine. In Proceedings of ICANN, 2001.
[666] Seidenberg M S. Language acquisition and Use: Learning and applying probabilistic constraints. Science, 1997, 275: 1599-1603.
[667] Sejnowski T. Time for a new neural code? Nature, 1995, 376: 21-22.
[668] Sellmann M. Intelligent Decisions. Invited Speaker, 29th AAAI Coference on Artificial Intelligence,

Austin, Texas, USA, January 24-30, 2015.

[669] Sewell W, Shah V. Social class, parental encouragement, and educational aspirations. American Journal of Sociology, 1968, 73: 559-572.

[670] Shafer G. A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press, 1976.

[671] Shannon C E. A mathematical theory of communication. Bell System Technical Journal, July and October, 1948, vol. 27: 379-423 and 623-656.

[672] Shapiro E Y. Inductive inference of theories from facts. Research Report 192, Department of Computer Science, Yale University, New Haven, CT. , 1981.

[673] Shareef N, Wang D L, Yagel R. Segmentation of medical images using LEGION. IEEE Transactions on Medical Imaging, 1999, 18: 74-91.

[674] Shaw R D, Phuoc T V, Paul A O, et al. Antigenic mapping of the surface proteins of rhesus rotavirus. Virology, 1986, 155: 434-451.

[675] Sherbondy A J, Dougherty R F, Ananthanaraynan R. et al. Think global, act local: Projectome estimation with BlueMatter. In Proceedings of the Medical Image Computing and Computer Assisted Intervention Society, Lecture Notes in Computer Science (London, Sept. 20-24). Springer, Berlin, 2009, 861-868.

[676] Sherrington C S. The Integrative Action of the Nervous System, 1961 Ed. Yale University Press, New Haven, CT. , 1906.

[677] Shi Zhongzhi. Distributed Artificial Intelligence, Future Directions in Artificial Intelligence, North-Holland, 1991.

[678] Shi Zhongzhi. Principles of Machine Learning. International Academic Publishers, 1992.

[679] Shi Zhongzhi. (Ed. ) Automated Reasoning. IFIP Transactions A-19, North-Holland, 1992.

[680] Shi Zhongzhi. (Ed. ) Proceedings of Pacific Rim International Conference on Artificial Intelligence, International Academic Publishers, 1994.

[681] Shi Zhongzhi. On Intelligence Science and Recent Progresses. IEEE ICCI 2006, pp. 16, 2006.

[682] Shi Zhongzhi. Progress of Artificial Intelligence. ICAI2006, 2006, 861-866.

[683] Shi Zhongzhi. On Intelligence Science. International Journal on Advanced Intelligence, 2009, 1(1): 39-57.

[684] Shi Zhongzhi. Intelligent Science. RSFDGrC2009: 20-32, 2009.

[685] Shi Zhongzhi. Research on Brain-Like Computer. Brain Informatics-2009: 5, 2009.

[686] Shi Zhongzhi. Image Semantic Analysis and Understanding. Intelligent Information Processing 2010: 4-5, 2010.

[687] Shi Zhongzhi. Advanced Artificial Intelligence. World Scientific Publishing Co. , 2011.

[688] Shi Zhongzhi. Foundations of Intelligence Science. International Journal of Intelligence Science, 2011, 1(1): 8-16.

[689] Shi Zhongzhi. Intelligence Science. World Scientific Publishing Co. , 2012.

[690] Shi Zhongzhi. Computational Model of Memory in CAM. Keynote Speaker, AGI-13, Workshop on Formal MAGIC, 2013.

[691] Shi Zhongzhi. Intelligence Science Is The Road To Human-Level Artificial Intelligence. Keynotes Speaker, IJCAI-13, Workshop on Intelligence Science, 2013.

[692] Shi Zhongzhi. Mind Computation. World Scientific Publishing Co. , 2017.

[693] Shi Zhongzhi, Faltings B, Musen M, et al Proceedings of Conference on Intelligent Information Processing. Publishing House of Electronics Industry, 2000.

[694] Shi Zhongzhi, Han Jianchao. Attribute Theory in Learning System, Future Generation Computer Systems, 1990, No. 6.

[695] Shi Zhongzhi, He Qing, Jia Ziyan, et al. Intelligence Chinese Document Semantic Indexing System. International Journal of Information Technology and Decision Making, 2003, 2(3): 407-424.

[696] Shi Zhongzhi, He Huang, Jiewen Luo, et al. Agent-based Grid Computing. Applied Mathematical Modeling, 2006, 30: 629-640.

[697] Shi Zhongzhi, Huang Youping, He Qing, et al. MSMiner-A Developing Platform for OLAP. Decision Support Systems, 2006, 42(4): 2016-2028.

[698] Shi Zhongzhi, Jiao Wenpin, Wang Bo. A Composition Model for Distributed Systems. ICCS2001, 2001, 590-594.

[699] Shi Zhongzhi, Luo Xudong. NLPM: An Approach for Nonmonotonic Reasoning Based on Linguistic Probability. IWAR 1992: 107-122.

[700] Shi Zhongzhi, Ma Gang, Yang Xi, et al. Motivation Learning In Mind Model CAM. International Journal of Intelligence Science, 2015, 5(2): 63-71.

[701] Shi Zhongzhi, Wu Bin, He Qing, et al. IDSIS: Intelligent Document Semantic Indexing System. Intelligent Information Processing 2002: 307-310.

[702] Shi Zhongzhi, Meng Zuqiang, Lu Yuan. Tolerance Granular Computing based on incomplete information system. GrC 2009: 501-506, 2009.

[703] Shi Zhongzhi, Shi Jun. Perspectives On Cognitive Informatics, IEEE ICCI'03, 129-136, 2003.

[704] Shi Zhongzhi, Tian Qijia, Li Yunfeng. RAO logic for multiagent framework. J. Comput. Sci. Technol. 1999, 14(4): 393-400.

[705] Shi Zhongzhi, Wang Xiaofeng. A Mind Model CAM: Consciousness and Memory Model. Proceedings of Cognitive Science, ICCS 2010: 148-149, 2010.

[706] Shi Zhongzhi, Wang Xiaofeng, Shi Zhiping, et al. A mind model for brain-like computer. IEEE ICCI 2010: 257-264, 2010.

[707] Shi Zhongzhi, Wang Xiaofeng. A Mind Model CAM In Intelligence Science. Progress of Advanced Intelligence, 2011, Vol. 2: 20-27.

[708] Shi Zhongzhi, Wang Xiaofeng. A Mind Model CAM in Intelligence Science. International Journal of Advanced Intelligence. 2011, 3(1): 119-129.

[709] Shi Zhongzhi, Wang Xiaofeng, Yue Jinpeng. Cognitive Cycle in Mind Model CAM. International Journal of Intelligence Science, 2011, 1(2): 25-34.

[710] Shi Zhongzhi, Zhang Jian, Liu Jimin. Neural Field Theory-A Framework of neural Information Processing. Proceedings of International Conference on Neural Network and Brain, 421-424, 1998.

[711] Shi Zhongzhi, Zhang Jianhua, Zhao Xiaofei. Motivation Model for Brain-Machine Integration. Technical Report, Institute of Computing Technology, Chines Academy of Sciences, 2012.

[712] Shi Zhongzhi, Zhang Jianhua, Yang Xi, et al. Computational Cognitive Models for Brain-Machine Collaborations. IEEE Intelligent Systems, Nov. /Dec. , 2014, 24-31.

[713] Shi Zhongzhi, Zhang Jianhua, Yue Jinpeng, et al. A Cognitive Model for Multi-Agent Collaboration. International Journal of Intelligence Science, 2014, 4(1): 1-6.

[714] Shi Zhongzhi, Zhang Sulan. Case-based Introspective Learning. IEEE ICCI'05, 2005.

[715] Shi Zhongzhi, Zheng Zheng, Meng Zuqiang. Image Segmentation-oriented Tolerance Granular Computing Model. GrC 2008: 566-571.

[716] Shi Zhongzhi, Zhou Han, Wang Jun. Applying case-based reasoning to engine oil design. AI in

Engineering,1997,11(2): 167-172.

[717] Shieber S.(Ed.) The Turing Test: Verbal Behavior as the Hallmark of Intelligence. The MIT Press,2004.

[718] Silver D, Hubert T, Hassabis D. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. arXiv: 1712.01815 [cs. AI],5 Dec 2017.

[719] Simon H A. Theories of Decision-Making in Economics and Behavioral Science. American Economic Review,1959,49 (1): 253-283.

[720] Simon H A. Information Processing Models of Cognition. Annual Review of Psychology,1979,30: 363-396.

[721] Simon H A. Models of Thought. New Haven,CT: Yale University Press,1979.

[722] Simon H A 著,荆其诚,张厚粲译.人类的认知:思维的信息加工理论.北京:科学出版社,1986.

[723] Simon H A. The Sciences of the Artificial(2ed). The MIT Press,1982.武夷山译.人工科学.北京:商务印书馆出版,1987.

[724] Singer W. Time as coding space in neocortical processing: a hypothesis. In: Buzsaki,G.,Lllinas R, Singer,W. et al. (eds),Temporal Coding in the Brain. Berlin: Springer-Verlag,1994,51-79.

[725] Skarda C A,Freeman W J. How brains make chaos in order to make sense of the world,Behavioral and Brain Sciences,1987,10: 161-195.

[726] Skinner B F. The science of learning and the art of teaching. Harvard Educational Review,1954,24 (2): 86-97.

[727] Sloman A. Varieties of Affect and the CogAff Architecture Schema. AISB2001,York,2001.

[728] Smale S. Complexity Theory and Numerical Analysis. Acta Numerica,1997,pp. 523-551.

[729] Smith E E,Shoben E J,Rips L J. Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review,1974,81: 214-241.

[730] Smith D J,Forrest S,Perelson A S. Immunological memory is associative. In; Dasgupta ed. Artificial Immune Systems and their Applications. Berlin: Springer,1998,105-112.

[731] Snider G,Rick A,Dick C,et al. From Synapses to Circuitry: Using Memristive Memory to Explore the Electronic Brain. IEEE Computer,2011,44(2): 21-28.

[732] Snaider J, McCall R, Franklin S. Time production and representation in a conceptual and computational cognitive model. Cognitive Systems Research,2012,13(1): 59-71.

[733] Sompolinsky H. Population Coding in Neuronal Systems with Correlated Noise[J]. Physics Review E,2001,64: 1-9.

[734] Song S,Miller K D, Abbott L F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat. Neurosci.,2000,. 3: 919-926.

[735] Sperry R W. Brain bisection and mechanisms of consciousness. In: Eccles JC (ed) Brain and Conscious Experience. New York: Springer-Verlag,1966,298-313.

[736] Sperry R W. Hemisphere deconnection and unity in conscious awareness. Scientific American,1968, 23: 723-733.

[737] Squire L R. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review99,1992: 195-232.

[738] Stein F,Medioni G. Map-based localization using the panoramic horizon[A]. Proceedings of the IEEE International Conference on Robotics and Automation[C]. Nice,France,1992,2631-2637.

[739] Sterling P R. in The Synoptic Organization of the Brain,G. M. Shepherd,ed.,3rd edition,pp. 170-213,New York: Oxford University Press,1990.

[740] Stojanov G. Petita96: A Case Study in Developmental Robotics. Proceedings of the first Intemational Workshop on epigenetic robotics, Lund, 2001.

[741] Stuart G J, Sakmann B. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. Nature, 1994, 367: 69-72.

[742] Suchman L A. Plans and situated actions: The problem of human-machine communication. Cambridge University Press, 1987.

[743] Sutter E E. The brain response interface: communication through electrical brain response. Microcomputer Appl, 1992, 15: 31-45.

[744] Sutton R S, Barto A G. An adaptive network that constructs and uses an internal model of its world. Cognition and Brain Theory, 1981, 3: 217-246.

[745] Sutton R S. Generalization in reinforcement learning: Successful examples using sparse coarse coding. In Touretzky D S, Mozer M C, and Hasselmo M E, (editors), Advances in Neural Information Processing Systems: Proceedings of the 1995 Conference, Cambridge, MA. MIT Press, 1038-1044, 1996.

[746] Sutton R S, Barto A G. Reinforcement Learning: An Introduction. MIT Press, 1998.

[747] Sutton R S. (editor) A Special Issue of Machine Learning on Reinforcement Learning, volume 8. Machine Learning. Also published as Reinforcement Learnng, Kluwer Academic Press, Boston, MA, 1992.

[748] Sutton R S. Temporal Credit Assignment in Reinforcement Learning. PhD thesis, University of Massachusetts, Amherst, MA, 1984.

[749] Sutton R S. Learning to predict by the method of temporal differences. Machine Learning, 1988, 3: 9-44.

[750] Sutton R S. Planning by incremental dynamic programming. In Birnbaum, L. A. and Collins, G. C., editors, Proceedings of the Eighth International Workshop on Machine Learning, Morgan Kaufmann, San Mateo, 1991, CA. 353-357.

[751] Sutton S, Bramn M, Zabin J, et al. Informalion delivery and the sensory evoked potential[J]. Science, 1965, 155: 1436-1439.

[752] Tan M. Multi-agent reinforcement learning: Independent vs. cooperative agents. InProceedings of the Tenth International Conference on Machine Learning, Morgan Kaufmann., 1993, 330-337.

[753] Terman L M, Merrill M A. Measuring Intelligence. Boston: Houghton Mifflin, 1937.

[754] Thagard P. Mind: Introduction to Cognitive Science. 1996, 2005. MIT Press. 保罗·萨伽德著. 心智:认知科学导论. 朱菁,陈梦雅译,上海:上海辞书出版社,2012.

[755] Thelen E., Smith L. B. A dynamic systems approach to the development of cognition and action. Cambridge, MIT Press, 1994.

[756] Tikhonov A N. Solution of incorrectly formulated problems and the regularization method. Soviet Math. Dokl., 1963, 4: 1035-1038.

[757] Timmis J, Knight T. Artificial immunes system: Using the immune system as inspiration for data mining. In: Abbass H A, Sarker R A, Newton C S eds. Data Mining: A Heuristic Approach. Hershey: Idea Publishing Group, 209-230, 2001.

[758] Tononi G. Consciousness as integrated information: A provisional manifesto. Biol. Bull. 2008, 215, 216-242.

[759] Tononi G., Sporns O. Measuring Integrated Information. BMC Neurosci., 2003, 4(31): 20.

[760] Toyoizumi Taro, Aihara Kazuyuki, Amari Shun-ichi. Fisher Information for Spike-Based Population Decoding[J]. Phys. Rev. L, 2006, 97: 1-4.

[761] Trabasso T, Bower G. Attention in Learning. New York: Wiley, 1968.
[762] Treisman A. Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 1960, 12: 242-248.
[763] Treisman A. Perceptual grouping and attention in visual search for features and for objects. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 194-214.
[764] Treisman A. The binding problem. Current Opinion in Neurobiology, 1996, 6: 171-178.
[765] Treisman A. Feature binding, attention and object perception. Philosophical Treansanctions of the Royal Society, Series B., 1998, 353: 1295-1306.
[766] Treisman A, Gormican S. Feature analysis in early vision: Evidence from search asymmetries. Psychological Review, 1988, 95: 15-48.
[767] Trueswell J C, Tanenhaus M K, Garnsey S M. Semantic influences on parsing: Use of Thematic role information in syntactic disambiguation. Journal of Memory and Language, 1994, 33: 285-318.
[768] Tsodyks M. Spike-timing-dependent synaptic plasticity-the long road towards understanding neuronal mechanisms of learning and memory. Trends Neurosci., 2002, 25: 599-600.
[769] Tulving E. Cue dependent forgetting. American science, 1974, 62: 74-82.
[770] Tulving E. Elements of episodic memory. London, Oxford Clarendon Press, 1983.
[771] Turing A M. On computable numbers with an application to the Entscheidungsproblem. Proc. London Maths. Soc., 1936 ser. 2, 42, 230-265.
[772] Turing A M. Computing Machinery and Intelligence. Mind, 1950, 59: 433-460.
[773] Tversky A. Choice by Elimination. Journal of Mathematical Psychology, 1972, 9: 341-367.
[774] Valiant L G. A theory of the learnable. Communications of the ACM, 1984, 27(11): 1134-1142.
[775] van der Meulen F F, Meyer A S, Levelt W J M. Eye movements during the production of nouns and pronouns. Memory and Cognition, 2001, 29: 512-521.
[776] van Essen D C.. Functional organization of primate visual cortex. In: Cerebral Cortex, Vol. 3, Ed. Peters A. and Jones EG. NewYoA: Plenum, 259-329, 1985.
[777] van Essen D. The Human Connectome Project: Progress and Perspectives. Keynotes Speaker, IJCAI-13 Workshop on Intelligence Science, http://www.intsci.ac.cn/WIS2013/, 2013.
[778] van Gelder T, Port R. It's about time: An overview of the dynamical approach to cognition. Mind as motion: Explorations in the dynamics of cognition. R. Port and T. van Gelder. Cambridge, MA, MIT Press, 1995.
[779] van Steveninck de Ruyter, Lewen R, Strong G, et al. Reproducibility and variability in neural spike trains[J]. Science, 1997, 275: 1805-1808.
[780] Vapnik V N. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995.
[781] Vapnik V N. Statistical Learning Theory. Wiley-Interscience Publication, John Wiley & Sons, Inc., 1998.
[782] Vapnik V N, Chervonenkis A Ja. The necessary and sufficient conditions for for consistency of the method of empirical risk minimization. Yearbook of the Academy of Sciences of the USSR on Recognition, Classification, and Forecasting, 1989.
[783] Vapnik V, Golowich S, Smola A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, in Advances in Neural Information Processing Systems 9, 1997.
[784] Vaseghi S V. State duration modeling in hidden Markov models. Signal Processing, 1995, 41: 31-41.
[785] Versace M, Chandler B. The Brain of A New Machine. IEEE Spectrum, Dec. 2010, 28-35. 史忠植译. 新型机器脑. 中国计算机学会通讯, 2011, 7(9): 70-76.

[786] Victor J. Binless strategies for estimation of information from neural data[J]. Physical Review E, 2002,66: 051903 1-051903 15.
[787] Vidal J J. Toward direct brain—computer communication. Annu Rev Biophv. 1973: 157-180.
[788] Voisin J, Devijver P A. An application of the Multiedit-Condensing technique to the reference selection problem in a print recognition system. Pattern Recognition,1987,5: 465-474.
[789] von der Malsburg C. How are nervous systems organized? In: Basar E, Haken H, Mandell A J eds. Synergetic of the Brain. Berlin: Springer,1983,238-249.
[790] von der Malsburg C. Nervous structures with dynamical links. Ber Bunsenges Phys Chem,1985,89: 703-710.
[791] von der Malsburg C. Binding in models of perception and brain function. Current Opinion in Neurobiology,1995,5: 520-526.
[792] von der Malsburg C,Schneider W A. Neural cocktail-party processor. Biol Cybern,1986,54: 29-40.
[793] Von Wright G H. (冯·赖特)胡泽洪,胡浩译. 二十世纪的逻辑和哲学. 哲学译丛,02,2000.
[794] Vygotsky L S. Mind in Society: The development of higher psychological processes. Cambridge, MA: Harvard University Press,1978.
[795] Vygotsky L S. Thought and language[M]. Cambridge: MIT Press,1986.
[796] Wainer H. Visual Revelations. Copernicus,1997.
[797] Waldrop M M. Computer modelling: Brain in a box. Nature,2012,482: 456-458.
[798] Watanabe W, Nakanishi K, Aihara K. Solving the binding problem of the brain withbidirectional functional connectivity. Neural Networks,2001,14: 395-406.
[799] Waterman M S,Perlwitz M D. Line Geometries for Sequence Comparisons. Bull Math Biol,1984,46 (4): 567-577.
[800] Watson John B. Psychology as the behaviorist views it. Psychological Review,1913,20: 158-177. http://psychclassics. yorku. ca/Watson/views. htm.
[801] Weng Juyang. Learning in image analysis and beyond: Development. Visual Communication and Image Processing,New York,1998.
[802] Weng Juyang. Autonomous Mental Development by Robots and Animals. Science,Issue of 26 Jan, 2001,291(5504): 599-600.
[803] Weng Juyang. Developmental Robotics: Theory and Experiments. International Journal of Humanoid Robotics,2004,1(2): 199-236.
[804] Weng Juyang. A 5-chunk developmental brain-mind network model for multiple events in complex backgrounds. IJCNN 2010: 1-8,2010.
[805] Werbos P J. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph. D. Thesis,Harvard University,Cambridge,MA. ,1974.
[806] Werbos P J. Backpropagation and neurocontrol: A review and prospectus. International Joint Conference on Neural Networks,vol. I,pp. 209-216,Washington,DC. ,1989.
[807] Wertheimer M. Laws of organization in perceptual forms. First published as Untersuchungen zur Lehre von der Gestalt Ⅱ,inPsycologische Forschung,1923,4: 301-350 Translation published in Ellis, W. A source book of Gestalt psychology (pp. 71-88). London: Routledge & Kegan Paul,1938.
[808] Wetmore D, Baker S. Post-spike distance-to-threshold trajectories of neurones in monkey motor cortex[J]. Journal of Physiology,2004,555: 831-850.
[809] Widrow B. Generalization and Information Storage in Networks of Adaline "Neurons". in Self-

Organizing Systems 1962, M. C. Yovitz, G. T. Jacobi, and G. Goldstein, eds. , Spartan Books, Washington, DC, 1962, pp. 435-461.
[810] Wiener N. Cybernetics, or control and communication in the animal and the machine. Cambridge, Massachusetts: The Technology Press; New York: John Wiley & Sons, Inc. , 1948.
[811] Williams J M G, Watts F N, MacLeod C M, et al. Cognitive psychology and emotional disorders. Chichester: John Wiley, 1988-1997.
[812] Wilson D. Asymptotic Properties of Nearest Neighbor Rules using Edited Data. Institute of Electrical and Electronic Engineers Transactions on Systems, Man and Cybernetics, 1972, 2: 408-421.
[813] Winograd T. Understanding Natural Language. Academic Press, 1972.
[814] Winograd T. Language as a Cognitive Process: Syntax, Addison-Wesley, 1983.
[815] Winograd T, and Fernando F. Understanding Computers And Cognition: A New Foundation for Design. Norwood, New Jersey: Ablex Publishing Corporation, 1986.
[816] Wolpaw J R, et al. An EEG-based brain-computer interface for cursor control. Electroenceph Clin Neurophysiol, 1991, 78: 252-259.
[817] Wolpaw J R, et al. Multichannel EEG-based brain-computer communication. Electroenceph Clin Neurophysiol, 1994, 90: 444-449.
[818] Woods W A. Transition Network Grammars for Natural Language Analysis, Comm. ACM, October 1970, 13, No. 10.
[819] Woods W A, Schmolze James. The KL-ONE Family. Computers & Mathematics with Applications, 1992, 23(1-5): 133-177.
[820] Wu S, Amari S, Nakahara H. Population Coding and Decoding in a Neural Field: A Computational Study. Neural Computation, 14 (5) May 2002 p. 999-1026.
[821] Xu Lei. Ying-Yang Machine: A Bayesian Kullback Scheme for Unified Learnings and New Results on Vector Quantization. The Proceeding of ICONIP'95, 1995, 977-988.
[822] Yachida M, Kitamura Y, Kimachi M. Trinocular vision: New approach for correspondenceproblem [A]. Proceedings of 8th Intentional Conference on Pattern Recognition[C]. Paris, France, 1986, 1041-1044.
[823] Yanger R R, Fedrizzi M, Kacprzyk J. Advances in the Dempster-Shafer Theory of Evidence. New York: Wiley, 1994.
[824] Yao Xin. Evolutionary artificial neural networks. International Journal of Neural Systems, 1993, 4 (3): 203-222.
[825] Yao Xin. Evolving artificial neural networks. Proceedings of the IEEE, 1999, 87(9): 1423-1447.
[826] Zadeh L A. Fuzzy Sets. Information and Control, 1965, 8: 338-353.
[827] ZadehL A. Fuzzy sets and information granularity. In: Gupta M M, Ragade R K, and Yager R R, eds. Advances in Fuzzy Set Theory and Applications, North-Holand, New York, 1979: 2-18.
[828] ZadehL A. Fuzzy Logic=Computing With Words, IEEE Trans. On Fuzzy Systems, 1996, 4: 103-111.
[829] Zajonc R B. Feeling and thinking: Preferences need no inferences. American Psychologist, 1980, 35: 151-175.
[830] Zajonc R B. On the primacy of affect. American Psychologist, 1984, 39: 117-23.
[831] Zemel R S, Dayan P, Pouget A. Probabilistie Interpretation of Population Codes [J]. Neural Computation, 1998, 10: 403-430.
[832] Zhang B, Zhang L. Theory and Application of Problem Solving. Elsevier Science Publishers, North-Holland, 1992.
[833] Zhang Jian, Shi Zhongzhi. An Adaptive Theoretical Foundation Toward Neural Information

Processing NFT, Proceeding of ICONOP'95, 1995, 217-220.
[834] Zhang Jian, Shi Zhongzhi, Liu Jimin. Topology Approximation Correction Approach—A Learning Mechanism of Neural Field Theory. Proceedings of International Conference on Neural Network and Brain, 1998, 421-424.
[835] Zhang Kechen, Ginzburg I, McNaughton B L, et al. Interpreting Neuronal Population Activity by Reconstruction: Unified Framework with Application to Hippocampal Cells[J]. Neurophysiology, 1998, 79: 1017-1044.
[836] Zhang Ling, Zhang Bo. A Geometrical Representation of McCulloch-Pitts Neural Model and Its Application, IEEE Transactions on Neural Networks, 1999, 10(4): 925-929.
[837] Zhang S, Du Z, Zhang L, et al. Cambricon-X: An accelerator for sparse neural networks. In: Proceedings of the 49th IEEE/ACM International Symposium on Microarchitecture (MICRO), Taipei, 2016.
[838] Zhang W. Reinforcement Learning for Job-shop Scheduling. PhD thesis, Oregon State University. Tech Report CS-96-30-1, 1996.
[839] Zhang W, Dietterich T G. A reinforcement learning approach to job-shop scheduling. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, 1114-1120.
[840] Zhang Ying, Brown R D, Frederking R E, et al. Pre-processing of Bilingual Corpora for Mandarin-English EBMT. In Proceedings of the MT Summit VIII. Santiago de Compostela, Spain, September, 2001.
[841] Zhang Ying, Shi Zhongzhi, You Xiangtao. A Decision Trees Learning Algorithm Based On Bias Shift. Proc. Knowledge Discovery and Data Mining Workshop, PRICAI'98, Singapore, 71-81, Nov., 1998.
[842] Zhou Z-H. Ensemble Methods: Foundations and Algorithms. Boca Raton, FL: Chapman & Hall/CRC, 2012.
[843] Zhuo Y, Zhou T G, Rao H Y, et al. Contributions of the visual ventral pathway to long-range apparent motion. Science, 2003, 299: 417-420.
[844] 白学军. 智力心理学的研究进展[M]. 杭州: 浙江人民出版社, 1997.
[845] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.
[846] 波珀. 科学发现的逻辑[M]. 查汝强, 邱仁宗, 译. 沈阳: 沈阳出版社, 1999.
[847] 波普尔. 猜想与反驳[M]. 傅季重, 等译. 上海: 上海译文出版社, 2001.
[848] 波普尔. 客观知识[M]. 舒伟光, 等译. 上海: 上海译文出版社, 2001.
[849] 蔡自兴. 基于功能/行为集成的自主式移动机器人进化控制体系结构[J]. 机器人, 2000, 22(3): 170-175.
[850] 曹虎. 多主体环境 MAGE 及其协作与规划研究[D]. 北京: 中国科学院计算技术研究所, 2000.
[851] 常亮. 扩展的动态描述逻辑及其应用研究[D]. 北京: 中国科学院计算技术研究所, 2008.
[852] 陈国良, 王煦法, 庄镇泉等. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 1996.
[853] 陈霖. 拓扑性质检测. 钱学森主编. 关于思维科学. 上海: 上海人民出版社, 1986.
[854] 陈晓平. 关于归纳逻辑的若干问题——对现代归纳逻辑的回顾与展望[J]. 自然辩证法通讯, 2000.
[855] 陈英和. 认知发展心理学[M]. 杭州: 浙江人民出版社, 1997.
[856] 戴汝为. 社会智能科学. 上海: 上海交通大学出版社, 2007.
[857] 丁世飞. 基于感知学习和语言认知的智能计算模型研究[D]. 北京: 中国科学院计算技术研究所, 2006.
[858] 董明楷. 面向智能主体的动态描述逻辑研究[D]. 北京: 中国科学院计算技术研究所, 2003.
[859] 董振东, 董强, 郝长伶. 知网的理论发现[J]. 中文信息学报, 2007, 21(4): 3-9.

[860] 冯志伟. 数理语言学[M]. 北京：知识出版社，1985.
[861] 耿海燕，朱滢. 关于无意识知觉的研究[J]. 心理学动态，1997，6(1)：14.
[862] 宫秀军. 贝叶斯学习理论及其应用研究[B]. 北京：中国科学院计算技术研究所，2002.
[863] 宫秀军，史忠植. 基于 Bayes 潜在语义模型的半监督文本挖掘[R]. 中国科学院计算技术研究所，2001.
[864] 桂起权，等. 次协调逻辑与人工智能[M]. 武汉：武汉大学出版社，2002.
[865] 郭爱克. 计算神经科学[M]. 上海：上海科技教育出版社，2000.
[866] 何华灿，等. 泛逻辑学原理[M]. 北京：科学出版社，2001.
[867] 何克抗. 创造性思维论——DC 模型的建构与论证. http://www.iteonline.net/zhuanjiaxuzhe/hekk.files/hekekang.htm.
[868] 何新贵，梁久祯. 过程神经元网络的若干理论问题[J]. 中国工程科学，2000，2(12)：40-44.
[869] 何新贵，许少华. 输入输出均为时变函数的过程神经网络及应用[J]. 软件学报，2003，14(04)：764-769.
[870] 胡世华，陆钟万. 数理逻辑基础[M]. 北京：科学出版社，1983.
[871] 黄昌宁. 统计语言模型能做什么[J]. 语言文字应用，2002.
[872] 黄瑞. Web 异构信息的语义表征与检索研究[D]. 北京：中国科学院计算技术研究所，2009.
[873] 黄友平. 贝叶斯网络研究[D]. 北京：中国科学院计算技术研究所，2005.
[874] 贾自艳. Web 信息智能获取若干关键问题研究[D]. 北京：中国科学院计算技术研究所，2004.
[875] 蒋运承. 基于主体的智能 Web 中的服务研究[D]. 北京：中国科学院计算技术研究所，2004.
[876] 焦文品. 基于进程演算的主体理论及开发方法研究[D]. 北京：中国科学院软件研究所，2000.
[877] 李其维. 论皮亚杰心理逻辑学. 上海：华东师范出版社，1990.
[878] 李清勇. 视觉感知的稀疏编码理论及其应用研究[D]. 北京：中国科学院软件研究所，2006.
[879] 李速，齐翔林，胡宏等. 功能柱结构神经网络模型中的同步振荡现象. 中国科学 C 辑，2004，34(4)：385-393.
[880] 李威. 移动主体的研究与应用. 北京：中国科学院计算技术研究所，1998.
[881] 李夏. 综合集成法和概念系统的结构. 博士学位论文，中国科学院自动化研究所，1998.
[882] 李晓黎，刘继敏，史忠植. 基于支持向量机与无监督聚类相结合的中文网页分类器. 计算机学报，2001，24(1)：62-68.
[883] 李晓黎. 网上信息检索与分类中的数据采掘研究[D]. 北京：中国科学院计算技术研究所，2001.
[884] 李衍达. 对智能研究的一些设想. 智能科学技术基础理论重大问题研讨会，国家自然科学基金委员会，2004.
[885] 李志欣. 图像语义标注和检索的研究[D]. 北京：中国科学院计算技术研究所，2010.
[886] 廖乐健. 约束满足问题的研究[D]. 北京：中国科学院计算技术研究所，1994.
[887] 林芬. 自主计算中自组织与自优化研究[D]. 北京：中国科学院计算技术研究所，2009.
[888] 梁吉业. 基于粗糙集与概念格的智能数据分析方法研究[D]. 北京：中国科学院计算技术研究所，2004.
[889] 梁永全. 基于内容的多媒体信息检索的研究[D]. 北京：中国科学院计算技术研究所，1999.
[890] 刘继敏. 基于形状图象检索的研究[D]. 北京：中国科学院计算技术研究所，2000.
[891] 刘奎林. 灵感发生论新探. 钱学森主编. 关于思维科学. 上海：上海人民出版社，1986.
[892] 刘曦. 图像理解中的物体识别与语义分析研究[D]. 北京：中国科学院计算技术研究所，2010.
[893] 刘晓力. 认知科学的动力学范式. 人工智能与哲学研讨会，厦门，2004.
[894] 刘亚，艾海舟，徐光佑. 基于主运动分析的野外视觉侦察系统——运动目标检测，跟踪及全景图的生成[J]. 机器人，2001，23(3)：250-256.

[895] 陆汝钤. 人工智能(上册,下册). 北京：科学出版社,1989,1996.
[896] 陆汝钤. 计算机语言的形式语义. 北京：科学出版社,1994.
[897] 路海东. 教育心理学. 长春：东北师范大学出版社,2002.
[898] 罗世维. 大规模人工神经网络理论基础. 北京：清华大学出版社,2004.
[899] 罗世维. 视觉信息认知计算理论. 北京：科学出版社,2004.
[900] 罗杰文. 主体网格中多主体协同研究[D]. 北京：中国科学院计算技术研究所,2008.
[901] 罗平. 分布式数据挖掘算法与信息理论的研究[D]. 北京：中国科学院计算技术研究所,2008.
[902] 马原野,王建红. 认知神经科学原理和方法. 重庆：重庆出版社,2003.
[903] 蒙祖强. 面向复杂信息系统的相容粒度计算方法及其应用研究[D]. 北京：中国科学院计算技术研究所,2009.
[904] 莫纯欢. 进化学习及人工生命中的突发行为[D]. 北京：中国科学院计算技术研究所,1996.
[905] 牛温佳. 基于语境的语义 Web 服务自动组合技术研究[D]. 北京：中国科学院计算技术研究所,2010.
[906] 潘谦红. 分布式信息检索的研究与应用[D]. 北京：中国科学院计算技术研究所,1999.
[907] 蒲慕明,徐波,谭铁牛. 脑科学与类脑研究概述. 中国科学院院刊,2016 年,31(7)：725-736.
[908] 钱学森. 开展思维科学的研究. 钱学森主编. 关于思维科学. 上海：上海人民出版社,1986.
[909] 邱莉榕. 基于主体的语义 Web 服务组合研究[D]. 北京：中国科学院计算技术研究所,2007.
[910] 沈炯. 汉语语调构造和语调类型. 方言,1994,No. 4.
[911] 施良方. 学习论著. 北京：人民教育出版社,1994.
[912] 史俊,史忠植. 论智能. 计算机科学,2005,32(6)：31-33,109-110.
[913] 石志伟. 感知的计算模型及其应用[D]. 北京：中国科学院计算技术研究所,2008.
[914] 史忠植. 知识工程. 北京：清华大学出版社,1988.
[915] 史忠植. 人类思维的层次模型. 特邀报告,中国人工智能联合会议,1990.
[916] 史忠植. 逻辑—对象知识模型. 计算机学报,1990,13(10)：787-791.
[917] 史忠植. 神经计算. 北京：电子工业出版社,1993.
[918] 史忠植. 高级人工智能(第一、二、三版). 北京：科学出版社,1998,2006,2011.
[919] 史忠植. 智能主体及其应用. 北京：科学出版社,2000.
[920] 史忠植. 人工智能研究进展. 第九届中国人工智能会议特邀报告,北京,2001.
[921] 史忠植. 高级计算机网络. 北京：电子工业出版社,2001.
[922] 史忠植. 知识发现(第一、二版). 北京：清华大学出版社,2002,2011.
[923] 史忠植. 展望智能科学. 科学中国人,2003,8：47-49.
[924] 史忠植. 意识系统模型. 汪云九,杨玉芳等编. 意识与大脑. 北京：人民出版社,226-231,2003.
[925] 史忠植. 基于感知学习和语言认知的智能计算模型. 技术报告,中国科学院计算技术研究所,2004.
[926] 史忠植. 智能科学. 北京：清华大学出版社,2006.
[927] 史忠植. 人工智能. 北京：国防工业出版社,2007.
[928] 史忠植. 认知科学. 北京：中国科技大学出版社,2008.
[929] 史忠植. 神经网络. 北京：高等教育出版社,2009.
[930] 史忠植. 智能科技创新驱动发展：从大数据到智能科学. 特邀报告,中国人工智能学会“创新驱动发展——大数据时代的人工智能”高峰论坛,深圳,2013. 10. 29.
[931] 史忠植. 心智计算. 北京：清华大学出版社,2015.
[932] 史忠植,常亮. 基于动态描述逻辑的语义 Web 服务推理. 计算机学报,2008,31(9)：1599-1611.
[933] 史忠植,董明楷,蒋运承等. 语义 Web 的逻辑基础. 中国科学 E 辑信息科学,2004,34(10)：1123-1138.
[934] 史忠植,蒋运承,张海俊等. 基于描述逻辑的主体服务匹配,计算机学报,2004,27(5)：625-635.

[935] 史忠植，林芬，罗杰文. 主体网格智能平台 AGrIP 构建及其应用. 智能系统学报，2006 1(1)：17-23.
[936] 史忠植，莫纯欢. 人工生命. 计算机研究与发展，1995，32(12)：1-9.
[937] 史忠植，史俊，郑金华. 人工生命的智力问题研究. 涂序彦，尹怡欣主编：人工生命及应用. 北京：北京邮电大学出版社，27-32，2004.
[938] 史忠植，汪涛，王文杰. 基于多主体的并行工程系统. 全国 CIMS 会议，1996.
[939] 史忠植，怀进鹏，田启家. 人工智能进展. CJCAI'2001，北京：清华大学出版社，2001.
[940] 史忠植，叶施仁，李晓黎等. 数据采掘. 世纪之交的知识工程和知识科学学术研讨会，1999.
[941] 史忠植，余志华. 认知科学和计算机[M]. 北京：科普出版社，1990.
[942] 史忠植，张长水，邓立，等. 前言. 深度学习专辑[J]. 计算机研究与发展，2014，51(9)：1889-1890.
[943] 史忠植，郑征. 相容粒度空间模型及其应用研究[M]. 北京：科学出版社，2007.
[944] 施智平. 大规模视频库的组织与检索技术研究[D]. 北京：中国科学院计算技术研究所，2005.
[945] 斯华龄. 电脑人脑化：神经网络——第六代计算机(普及本). 北京：北京大学出版社，1992.
[946] 陶建华，蔡莲红. 基于统计模型的韵律建模思路[J]. 中文信息学报，2002.
[947] 田启家，史忠植. 动作和进化的逻辑基础[J]. 中国科学(E 辑)，1997，27(3)：282-288.
[948] 童天湘. 智能化是信息化发展的必然趋势[C]. 智能科学技术基础理论重大问题研讨会，国家自然科学基金委员会，2004.
[949] 涂序彦，尹怡欣. 人工生命及应用[M]. 北京：北京邮电大学出版社，2004.
[950] 韦义平. 心理与教育研究数据处理技术[M]. 桂林：广西师范大学出版社，2002.
[951] 王净，杨玉芳. 听觉场景分析理论的评价[J]. 心理学动态，1998，6：1-5.
[952] 王伟. 基于内容的图象信息检索系统的研究[D]. 北京：中国科学院计算技术研究所，1999.
[953] 王晓峰. 面向视频理解的认知建模及其应用研究[D]. 北京：中国科学院计算技术研究所，2011.
[954] 王雨田. 现代逻辑科学导引[M]. 北京：中国人民大学出版社，1987.
[955] 王雨田，吴炳荣. 归纳逻辑与人工智能[M]. 上海：上海中国纺织大学出版社，1995.
[956] 汪云九，杨玉芳，等. 意识与大脑[M]. 北京：人民出版社，2003.
[957] 汪云九. 神经信息学——神经系统的理论和模型[M]. 北京：高等教育出版社，2006.
[958] 王竹晓. 面向分布式信息检索的自动推理[D]. 北京：中国科学院计算技术研究所，2010.
[959] 吴边，苏煜，张剑慧等. 基于 P300 电位的新型 BCI 中文输入虚拟键盘系统[J]. 电子学报，2009，37(8)：1733-1738，1745.
[960] 吴斌. 群体智能的研究及其在知识发现中的应用[D]. 北京：中国科学院计算技术研究所，2002.
[961] 武成岗. 移动主体部分关键技术研究[D]. 北京：中国科学院计算技术研究所，2001.
[962] 吴健辉，罗跃嘉. 情绪的认知科学研究途径[C]. 第一届中国情感计算及智能交互学术会议，2003.
[963] 吴文俊. 吴文俊论数学机械化[M]. 济南：山东教育出版社，1996.
[964] 吴朝晖，俞一鹏，潘纲等. 脑机融合系统综述[J]. 生命科学，2014，26(6)：645-649.
[965] 21 世纪 100 个科学难题编写组. 21 世纪 100 个科学难题[M]. 吉林：吉林人民出版社，1998.
[966] 席裕庚. 动态不确定环境下广义控制问题的预测控制[J]. 控制理论与应用，2000，17 (5)：665-670.
[967] 徐科. 神经生物学纲要[M]. 北京：科学出版社，2000.
[968] 杨锦陈，杨玉芳. 言语产生中的韵律生成[J]. 心理科学进展，2004，12(4)：481-488.
[969] 杨文俊. 大脑高级功能的神经电生理[M]. 北京：中国科技出版社，1998.
[970] 叶浩生. 西方心理学的历史与体系[M]. 北京：人民教育出版社，1998.
[971] 叶世伟. 前向神经网络变换研究[D]. 北京：中国科学院计算技术研究所，1995.
[972] 张钹，张铃. 问题求解的理论与应用[M]. 北京：清华大学出版社，1990.
[973] 张海俊. 基于主体的自主计算研究[D]. 北京：中国科学院计算技术研究所，2005.
[974] 张建，史忠植. 柔性信息处理的框架理论-神经场理论，人工智能新进展[M]. 北京：清华大学出版

社,1994.
[975] 张建,史忠植.多层随机神经网络 EM 算法.计算机研究与发展,1996,33(11):808-815.
[976] 张建.基于微分流形神经场计算理论及其在金融分析中的运用[D].北京:中国科学院计算技术研究所,1996.
[977] 张素兰.知觉组织模型及应用研究[D].北京:中国科学院计算技术研究所,2007.
[978] 张颖.数据采掘的研究与应用[D].北京:中国科学院计算技术研究所,1998.
[979] 赵松年,熊小芸,姚国正等.视—脑信息处理的新进展[J].中国科学 C,1999,42(5):1015-1026.
[980] 赵志崑.主体统一建模语言研究[D].北京:中国科学院研究生院,2003.
[981] 郑征.相容粒度空间模型及其应用研究[D].北京:中国科学院计算技术研究所,2006.
[982] 钟义信.高等人工智能原理[M].北京:科学出版社,2013.
[983] 周昌乐.心脑计算举要[M].北京:清华大学出版社,2003.
[984] 周涵.基于范例学习的内燃机油产品设计系统[D].北京:北京石油大学出版社,1993.
[985] 周志华.机器学习[M].北京:清华大学出版社,2016.

# 图 书 资 源 支 持

感谢您一直以来对清华版图书的支持和爱护。为了配合本书的使用，本书提供配套的资源，有需求的读者请扫描下方的“清华电子”微信公众号二维码，在图书专区下载，也可以拨打电话或发送电子邮件咨询。

如果您在使用本书的过程中遇到了什么问题，或者有相关图书出版计划，也请您发邮件告诉我们，以便我们更好地为您服务。


**我们的联系方式：**

地　　址：北京市海淀区双清路学研大厦 A 座 701

邮　　编：100084

电　　话：010－62770175－4608

资源下载：http://www.tup.com.cn

客服邮箱：tupjsj@vip.163.com

QQ：2301891038（请写明您的单位和姓名）


教学交流、课程交流



清华电子



扫一扫，获取最新目录

**用微信扫一扫右边的二维码，即可关注清华大学出版社公众号“清华电子”。**